DOCUMENT RESUME

ED 034 290	24	EA 002 594

AUTHOR	Bloom, Benjamin S.
TITLE	Cross-National Study of Educational Attainment: Stage I of the I.E.A. Investigation in Six Subject Areas. Final Report. Volume I.
INSTITUTION	Chicago Univ., Ill.; International Project for the Evaluation of Educational Achievement, Hamburg (West Germany).
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-6-2527
PUB DATE	Feb 69
GRANT	OEG-3-6-062527-2226
NOTE	386p.
EDRS PRICE	EDRS Price MF-$1.50 HC-$19.40
DESCRIPTORS	*Academic Achievement, Civics, Cognitive Tests, Cross Cultural Studies, Data Processing, *Educational Research, *Elementary Education, English (Second Language), Evaluation Techniques, French, Interagency Cooperation, *International Education, Literature, Measurement Instruments, Questioning Techniques, Reading Comprehension, Science Education, *Secondary Education
IDENTIFIERS	Australia, Belgium, Chile, England, Federal Republic of Germany, Finland, France, IEA, India, *International Evaluation Educational Achievement, Iran, Italy, Japan, Netherlands, New Zealand, Poland, Scotland, Sweden, Thailand, United States

ABSTRACT

Twelve chapters describe an 18-nation study of educational attainment at the elementary and secondary school levels. Professional researchers participated in the project, developed through UNESCO and conducted under the auspices of educational research centers in Australia, Belgium, Chile, England, the Federal Republic of Germany, Finland, France, India, Iran, Italy, Japan, the Netherlands, New Zealand, Poland, Scotland, Sweden, Thailand, and the United States. Subjects of the chapters include: Testing of achievement in science; reading comprehension; literature; French as a second language; English as a second language; civic education; development of measuring instruments; hypotheses, sampling designs, and between country analyses; plans for data collection; processing procedures; and consequences of IEA studies on education research and policy decisions. Ten appendices include proposals for future IEA work and basic testing rationales. (An expanded and revised version of Appendix F-1 appears as ED 018 431.) Related documents are Volume II, sections 1-4 and 6, (EA 002 615) and Volume II, section 5, (FL001 477). (JK)

BR-6-2527
PA-24
OE/BR

ED034290

FINAL REPORT

Project No. 6-2527

Grant No. HEW-OEG-3-6-062527-2226

CROSS NATIONAL-STUDY OF EDUCATIONAL ATTAINMENT:

STAGE I OF THE I.E.A. INVESTIGATION IN SIX SUBJECT AREAS

Volume I

Prepared by

Participants in the International Project
for the Evaluation of
Educational Achievement (I.E.A.)

Principal Investigator

Benjamin S. Bloom
The University of Chicago
Chicago, Illinois

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

February 1969

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

EA 002 594

VOLUME I

TABLE OF CONTENTS

Page

CHAPTER 1. INTRODUCTION AND HISTORY OF THE PROJECT

1.0 Introduction 1-1

1.1 Initiation of IEA 1-1

1.2 Pilot Study 1-2

1.3 Mathematics Study (Phase I) 1-3

1.4 A Six-Subject Area Study (Phase II - Stage I) 1-5

1.5 Formation of International and National Committees. . . 1-7

1.6 IEA Organization. 1-11

1.7 Participation in IEA. 1-11

CHAPTER 2. SCIENCE

2.0 Introduction. 2-1

2.1 First Stage 2-2

2.2 Second Stage (Topic and Objective Specifications) . . . 2-4

2.3 Third Stage (Pre-Testing) 2-5

2.4 Construction of the Final Tests 2-9

2.5 Student, Teacher and School Questionnaire 2-14

2.6 General Appraisal 2-14

CHAPTER 3. READING COMPREHENSION

3.0 Introduction. 3-1

3.1 Preparation of Rationale. 3-2

3.2 Preparation of Materials. 3-3

3.3 Pre-Testing. 3-3

3.4 Preparation of Final Tests. 3-4

3.5 Reading Speed 3-4

3.6 Word Knowledge Test 3-6

Page

CHAPTER 4. LITERATURE

4.0 Introduction. 4-1

4.1 Preparation of Rationale. 4-1

4.2 The Selection of the Goals. 4-2

4.3 Populations to be tested. 4-3

4.4 The Instruments to be Used. 4-3

4.5 Preparation of Tests. 4-4

CHAPTER 5. FRENCH AS A FOREIGN LANGUAGE

5.0 Introduction. 5-1

5.1 Preparation of Rationale. 5-1

5.2 Preparation of Materials. 5-2

5.3 Pre-Testing. 5-4

5.4 Preparation of Penultimate Versions of Tests. 5-8

5.5 Timetable for Future Activities of the Committee. . . 5-9

CHAPTER 6. ENGLISH AS A FOREIGN LANGUAGE

6.0 Introduction. 6-1

6.1 Preparation of Rationale. 6-1

6.2 Preparation of Materials - First Stage. 6-2

6.3 Pre-Testing. 6-6

6.4 Attitude and Other Descriptive Measures 6-7

6.5 Points of Principle. 6-7

6.6 Future Action 6-12

CHAPTER 7. CIVIC EDUCATION

7.0 Introduction 7-1

7.1 Contents of the National Documents. 7-2

7.2 The Development of Cognitive Instruments. 7-3

Page

7.3 The Civic Education Content Specifications. 7-3

7.4 The Civic Education Pretests. 7-4

7.5 Criteria for Selection of Items for Final Draft Tests . 7-6

7.6 The Development of Affective Instruments. 7-7

7.7 Interrelation Between Cognitive and Affective Domains . 7-13

7.8 Status of Instruments. 7-13

CHAPTER 8. DEVELOPMENT OF QUESTIONNAIRES AND ATTITUDE AND DESCRIPTIVE SCALES

8.0 Questionnaires. 8-1

8.1 Example of Pre-Testing in Chile 8-3

8.2 Review of Pre-Testing and Use of Unscaled Variables . . 8-4

8.3 Attitude and Descriptive Scales 8-8

8.4 General Attitude Scales. 8-9

8.5 Science Attitude Scales. 8-11

8.6 Literature Attitude Scales. 8-12

8.7 French Attitude Scales. 8-13

8.8 English as a Foreign Language Attitude Scales 8-14

8.9 Civic Education Attitudes. 8-14

8.10 Descriptive Scales 8-14

8.11 General Descriptive Scales 8-15

8.12 Science Descriptive Scales. 8-15

CHAPTER 9. HYPOTHESES, SAMPLING DESIGNS AND BETWEEN COUNTRY ANALYSES

9.0 Hypotheses 9-1

9.1 Sampling Designs 9-9

9.2 Analyses Between Countries. 9-12

Page

CHAPTER 10. PLANS FOR ADMINISTRATION AND DATA COLLECTION

10.0 Introduction. 10-1

10.1 Separation of Stage 2 and Stage 3 10-2

10.2 The Need for a Dry Run 10-2

10.3 The Three Administrative Manuals 10-4

10.4 The Development of the Administrative Procedures. . . 10-5

10.5 General Plans for Calendar and Writing Assignments. . 10-6

CHAPTER 11. THE DEVELOPMENT OF DATA PROCESSING PROCEDURES

11.0 Introduction. 11-1

11.1 Timetable 11-3

CHAPTER 12. POSSIBLE CONSEQUENCES OF IEA STUDIES ON EDUCATIONAL RESEARCH AND EDUCATIONAL POLICY DECISIONS 12-1

REFERENCES r-1

APPENDICES

Page

APPENDIX A PROPOSALS TO IEA FOR FUTURE WORK

A-1 Civics (Proposal for an International Investigation of Instruction in Civics). A-1

A-2 English (Testing the Ability in English as a Foreign Language). A-10

A-3 Mother Tongue (Possibilities of an International Research Project in Mother Tongue at the Secondary School Level). A-24

A-4 Literature (Proposal for an International Study of Student Responses to Literature) A-41

A-5 French (Report of the USA Committee on the Testing of French as a Foreign Language) A-53

APPENDIX B EXAMPLES OF NEWSLETTERS

B-1 Newsletter of June,1968 . . . . B-1

B-2 Newsletter of December,1968. . . . . B-9

APPENDIX C STATUTES C-1

APPENDIX D SCIENCE DOCUMENTS

D-1 First Grid of Topic Areas and Objectives in Science. . . . D-1

D-2 a) Table 1 - Expanded List of Teaching Objectives. D-5

b) Table 2 - List of Subject Areas D-7

D-3 Modification of Grid of Topics and Objectives D-13

APPENDIX E READING COMPREHENSION RATIONALE E-1

APPENDIX F LITERATURE DOCUMENTS

F-1 The Elements of Criticism and Interpretation. F-1

F-2 Literature Rationale F-13

F-3 Literature Cross-Over Study. F-21

Page

APPENDIX G FRENCH DOCUMENTS

G-1 French as a Foreign Language Rationale G-1

G-2 French Test Rationale and Description of Pre-Tests. G-9

APPENDIX H CIVIC EDUCATION DOCUMENTS

H-1 Conceptual Framework for Civic Education. H-1

H-2 Illustrative Items for Civic Education H-33

H-3 Report on the IEA Pilot Interviews for Civic Education (Affective Domains) by Margarete Harvey H-39

APPENDIX I SPECIFIC SUBJECT MATTER HYPOTHESES AND QUESTIONS I-1

APPENDIX J THE ANALYSIS AND PRESENTATION OF THE IEA EVIDENCE J-1

CHAPTER I

INTRODUCTION AND HISTORY OF THE PROJECT

1.0 Introduction

This report describes the development of the International Project for the Evaluation of Educational Achievement (IEA) in recent years. In particular, it describes in detail the work undertaken between November, 1966 and February, 1969. This period has been the first stage of the second major phase of IEA's work. During this stage, international groups have constructed and pretested cognitive and affective instruments to evaluate the outcomes of learning in six subject areas at three different points in the school systems in twenty countries. The first point in the school system comprises the 10 year old level, the second is the 14 year old level, and the third is the final year of secondary school or the pre-university level. In addition to the evaluation instruments, questionnaires have been constructed to include measures of variables, descriptive of the student and his home background, the teachers in school, the curriculum and instruction, and the school characteristics. Procedures for the data collection and data processing have been worked out for the field work in the period 1969-71.

Volume I of this report includes a description of various parts of this phase of the study accompanied by supporting documents which are to be found in the appendices. Volume II of the report is the collection of instruments and administrative manuals which have been produced in this phase of the IEA work.

1.1 Initiation of IEA

In the middle fifties, groups of educators and educational researchers from different countries met at the UNESCO Institute for Education, in Hamburg, Germany, to examine problems such as those concerned with school structures and organization, selection processes, examinations and failure in school. Two important publications emerging from some of these meetings were edited by Hotyat (1962) and Wall (1962). Throughout these meetings there was a growing awareness of the need to establish evaluation techniques which would be valid cross-nationally. At the same time, more or less independently of each other, several researchers in the United States (Anderson, Bloom, and Foshay) began to consider the possibilities of undertaking such research.

1.2 Pilot Study

In 1958, researchers from several countries came together at a meeting in Eltham, England, chaired by Dr. W.D. Wall of the National Foundation for Educational Research in England and Wales, and also at the UNESCO Institute for Education in Hamburg. At these meetings it was decided to carry out a pilot study to discover if an international research project would be administratively possible and if the results could be meaningful. Research Centers from Belgium, England, Finland, France, Germany, Israel, Poland, Scotland, Sweden, Switzerland, the United States and Yugoslavia took part. the target population was all children in these countries aged 13:0 - 13:11, since this was the last point where all of an age group were still in school in all these countries. In most cases children in schools or areas which were known to be near to the national mean and standard deviation were tested, and thus there was no strict probability sampling. All in all, 9,918 pupils spread over eight languages were administered tests of reading comprehension, mathematics, science, geography and nonverbal ability. Altogether there were 120 test items, and the tests were edited by culling from existing tests, for the most part, items thought to be appropriate. The results of this pilot study were reported in a book edited by Foshay (1962).

Despite the evident limitations of the pilot study in terms of sampling and test construction, some of the findings were very interesting in the sense that they were unexpected. It was found that when the variance between countries in a subject was expressed as a percent of the average within-country variance, then that for reading comprehension was 6 percent; mathematics, 15 percent, science, 5 percent; geography, 15 percent; and nonverbal ability, 12 percent. This was surprising, since it was expected by some that the international variance for nonverbal ability would be least.

Another interesting finding was that boys varied more from country to country than girls. Professor R.L. Thorndike drew up national profiles or patterns of achievement. A country's pattern of achievement showed that country's achievement on specific tests and subtests relative to its own overall level of achievement. The profiles indicated, for example, that French pupils performed relatively better in mathematics than in science, but that the reverse was true in Germany and the United States. Despite the shorter length of schooling of pupils in Scandinavia (pupils there begin school at the age of 7), the 13 year olds there performed, in general, as well on the tests as children in other countries with longer schooling.

Apart from supplying information which was extremely useful in the generation of hypotheses for the IEA project, the answer to the main problem of the pilot study was that it was possible and meaningful to carry out an international project of this kind on a fairly large scale. Work then began immediately on the first main phase of the IEA project.

1.3 Mathematics Study (Phase I)

In 1961, the heads of research organizations and university departments of education, in twelve countries, met together to form the Council of the International Project for the Evaluation of Educational Achievement (IEA). These countries were: Australia, Belgium, England, Finland, France, Germany, Israel, Japan, Netherlands, Scotland, Sweden and the United States.

The Council decided that the first phase of the project should consist of an examination of attainment in one subject area only - mathematics - together with the collection of data concerning the factors that are associated with performance.

From the outset it was agreed that the project should be a co-operative enterprise. Matters of policy have been determined by full meetings of the Council. Interim decisions have been taken by an elected Standing Committee, and the day-to-day administration has been the responsibility of the Chairman, the Technical Director and a full-time Co-ordinator.

The participating research center from each country made itself responsible for the administration of tests and the collection of data within its boundaries. In other respects resources were pooled in the sense that wherever particular facilities existed they were placed at the disposal of the Council. For example, the United States representatives were able to arrange for all the data processing to be carried out at the University of Chicago Computation Center; England supplied an expert in sampling to advise on all the sampling procedures; and the United States, Belgium, and Sweden supplied mathematicians to serve as consultants to the test constructors. The costs of the operation occurring within each country (e.g. in carrying out the testing program, etc.) were borne by the country concerned, and the international costs were defrayed by the United States Office of Education.

The major purpose of the inquiry was the measure achievement in mathematics, and to relate this to those factors in home, school and society which might have been expected to be associated with it. In determining these factors, the investigation had, of course, to lean heavily upon the findings of previous research carried out in the main within particular educational systems. The project represents an attempt to assess the "efficiency" or "productivity" of different educational systems and practices. The countries participating were the twelve mentioned above. Two major points in the school systems were sampled:

a) 13 year old (both age and grade populations), since this was the last point in all of the systems where one hundred percent of an age group were still in full-time schooling, and

b) the pre-university grade students.

In some countries students at a major terminal point intermediate to 13 year olds and pre-university students were also tested. Apart from the international mathematics tests being administered, an opinion booklet (description scales of the school and class learning environment as well as various attitude scales,) student questionnaires, mathematics teachers' questionaires, and national case study questionnaires were also administered. All in all, just over 133,000 students from 5,400 schools were tested and questionnaires were completed by 13,500 teachers and 5,400 school principals. The number of single pieces of data collected was just over 50 million.

It was possible to examine the "effectiveness" of different educational practices in school and class organization, curricula, performance. Furthermore, outcomes were related to social, economic and ecological factors. The hypotheses tested can be grouped under the following four main headings:

1. The relation of school organization to mathematics achievement.

2. Curriculum and instruction.

3. Support for education.

4. Social, economic and ecological factors.

The results have been published in Husén, (et.al.,1967).

Apart from many findings relevant to the improvement of mathematical education in each of the participating countries, it should be noted that it has been possible to construct a "fair" and valid international mathematics test and through it to establish national norms for several target populations on the international test. The implications of this for comparison of pre-university grade students seeking college entrance in countries other than their own are obvious.

A Data Bank (Wolf, 1967) has been established and is open to all bona fide scholars.

Most of the national research centers have prepared reports on the Mathematics Study in which they have highlighted the implications of this research for their own schools. By the end of 1968 national reports had been produced by the following research institutions: Australia, (Keeves, 1968); Belgium (De Landsheere, 1967); England (Pidgeon, 1967); Finland (Kuusinen, 1967, and Lyttkainen & Pavliinen, 1969); Japan (N.I.E.R., 1967), and Sweden (Hultin, 1968).

1.4 A Six Subject Area Study (Phase II - Stage 1)

Already in 1964 it was clear that it would be necessary to repeat the Mathematics Study in other subject areas. Therefore IEA asked various of the existing National Centers to carry out for them what were called "Feasibility Studies" in certain subject areas. National Centers set up committees who furnished IEA with reports on the possibilities and worthwhileness of conducting surveys in selected subject areas. It was evident that Science was a strong candidate for future study as a subject which would not be too different from mathematics. Therefore no feasibility study was carried out in this area. However, the Swedish National Center produced a report on Civic Education (see Appendix A-1), the Finnish National Center one on English as a Foreign Language (see Appendix A-2), the Belgian National Center one on Mother Tongue (see Appendix A-3), and a special U.S. Committee headed by Professors Foshay and Purves produced one on Literature (see Appendix A-4). Another U.S. Committee, which was headed by Professor John B. Carroll, then of Harvard University, supplied a report on French as a Foreign Language (see Appendix A-5).

These reports were considered by the IEA Council in February 1965, and it was decided that an attempt should be made to construct instruments in each of the subject areas. There were some doubts expressed by some members of the Council as to whether valid international instruments could be constructed, in some areas, especially in Literature and Civic Education.

Not only was it decided to attempt to repeat the Mathematics Project in these six other subject areas, but also to extend considerably the background information collected (the "inputs" into the school systems) to try to account for more of the variance of the "outcomes" of the school systems (in terms of the students' performance on the tests) and by so doing to lay the factual basis for decision making for school reform - both national (within nations) and international (comparability of "standards" etc. between nations).

Phase II was split into two stages initially; Stage 1 was for the construction of the instruments, and Stage 2 for the field work, the analyses, and the write-ups.

Stage 1 began in November 1966 and continued until March 1969. Although an attempt was made to construct instruments in all six subject areas, it was anticipated that IEA might not be successful in each one of these areas. It was agreed that the instruments produced in Stage 1 would be reviewed at the end of 1968 in order to decide on those subject areas which would continue into Stage 2 for the field testing. It was found necessary in fact to split Stage 2 into Stage 2 and Stage 3, as the burden of testing all six subject areas at the same time was excessive. In Stage 2, therefore, only Science, Reading Comprehension ane Literature will be field tested, and Stage 3 will consist of English as a Foreign Language, French as a Foreign Language and possibly Civic Education.

The main aims of Stage 1 have been the construction of the following instruments:

- Tests in each of the subject areas for 10 year olds, 14 year olds, and pre-university students, as well as for major terminal populations between 14+ and the pre-university students;

- Accompanying student, teacher, school and national questionnaires;

- Descriptive and attitude scales.

At the same time, sampling designs and proposals have been developed, hypotheses formulated and data analysis programs planned.

The nations participating in Stage 1 have been Australia, Belgium, Chile, England, Federal Republic of Germany, Finland, France, India, Iran, Italy, Japan, Netherlands, Poland, Scotland, Sweden, Thailand, United States, New Zealand.

1.5 Formation of International and National Committees

Once that IEA had decided to attempt to construct cognitive achievement tests in the six subject areas, the first step was to form international subject area committees, the members of which would be drawn from various countries and be knowledgeable in the field of test construction in a particular subject area. To this end all National Centers were circularized to propose at least one person for the subject areas in which they were participating and to send full information about that person. Once the information was received, it was reviewed by the Bureau, and in the first instance chairmen were selected, and then a selection made from other nominees such that different cultures and different experiences were represented within one committee. Once the names had been agreed with the chairmen, letters of invitation were sent to proposed members. Where a member declined the invitation because of too many work commitments, further discussion took place with the chairman of the committee and in some instances other nominations which had been made were taken into consideration and in other cases new names were suggested. The composition of committees was finalised by October 1966. In some cases members were not able to continue with their work, in which case other persons were put on the committee, and hence in the committees as they now stand, there are occasionally more than one representative of a nationality on one committee. In some cases as a member could not attend one meeting, the opportunity was taken to bring in another expert in the field for just that committee meeting. Pages 19-29 of Booklet 6 in Volume II present the names and biographical notes of the members of the International Committees.

National Centers were requested to form national committees in each of the subject areas in which they were participating. Again members were expected to have some knowledge of test construction, but at the same time the committee should be so composed as to represent the different types of schools within any one designated population within the country. Where National Centers had a person from their country on the International subject area committee, they were asked that this person should be chairman of the national subject area committee, or at least a member of it, since this person was to act as a link between the two committees.

How National Centers organized the national committees and the number of members on the committees was at their own discretion, and from page 91-99 of Booklet 6 it will be seen that in some countries there were many members, and in others only a few. In Booklet 6 the names and positions of all members of national subject area committees are presented.

As will be seen from the separate subject area reports below, there was a continuous exchange of ideas between the international and national subject area committees. The international committee decided on the rationale of the test construction and asked for comments on this from the national subject area committees. Content analyses were required from the national subject area committees, and this indeed was an enormous amount of work. Furthermore, national committees were requested to submit test items to the international committees. Once the pre-tests were in first draft form, the national committees were asked to comment on them and the pre-tests were revised in the light of these comments before being submitted for pre-testing in the countries. The national committees were also asked to comment on the pre-test item analysis from their countries to the international committee. The international committee proposed a final test version and national committees were again asked to comment on these items and the final versions were revised in the light of their comments. In many cases it was members of the national committees who were responsible for the translation of the items into their own language for testing purposes. How active national committees were depended to a great extent on the National Center, but all committees worked hard and much of the good quality of the final items is due to their dedication and professional interest. Table 1 shows the National Centers participating in the different subject areas in Stage 1.

TABLE I

STATE OF PRE-TESTING		Australia	Belgium	Chile	England	Fed.Rep. of Germany	Finland	France	India	Iran	Italy	Japan	Netherlands	Poland	Scotland	Sweden	Thailand	U.S.A.	New Zealand
Reading Comprehension	Pop.I	–	✓	✓	✓	–	✓	–	✓	✓	✓	–	✓	✓	✓	✓	✓	✓	–
	II	–	✓	✓	✓	–	✓	–	✓	✓	✓	–	✓	✓	✓	✓	✓	✓	–
	IV	–	✓	✓	✓	–	–	–	–	✓	✓	–	✓	✓	✓	✓	✓	✓	–
French as a Foreign Lang.	Pop.I	–	–	–	✓	–	–	–	–	–	–	–	–	–	✓	–	–	✓	–
	II	–	–	–	✓	–	–	–	–	✓	✓	–	✓	–	✓	–	–	✓	–
	IV	–	–	✓	✓	–	–	–	–	✓	✓	–	✓	–	✓	✓	–	✓	–
English as a Foreign Lang.	Pop.II	–	–	–	–	✓	✓	–	–	✓	✓	–	✓	–	–	✓	✓	–	–
	IV	–	–	✓	–	✓	✓	–	–	✓	✓	–	✓	–	–	✓	✓	–	–
Science	Pop.I	–	–	✓	✓	✓	✓	–	✓	✓	✓	✓	–	✓	–	✓	✓	✓	–
	II	✓	✓	✓	✓	✓	✓	–	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	IV	✓	✓	✓	✓	✓	–	✓	–	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	IVS	✓	–	✓	✓	✓	–	–	–	✓	✓	–	✓	–	✓	✓	✓	✓	✓
Civic Education	Pop.I	–	–	–	✓	✓	✓	–	–	✓	✓	–	–	–	–	✓	–	✓	–
	II	–	–	–	✓	✓	✓	–	–	✓	✓	–	–	–	–	✓	–	✓	–
	IV	–	–	✓	✓	✓	–	–	–	✓	✓	–	–	–	–	✓	–	✓	–
	IVS	–	–	–	✓	✓	–	–	–	✓	✓	–	–	–	–	✓	–	✓	✓
Literature: Mult. Choice	Pop.II	–	✓	✓	✓	–	✓	–	–	✓	✓	–	–	–	–	✓	–	✓	–
	IV	–	✓	✓	✓	–	–	–	–	✓	✓	–	–	–	–	✓	–	✓	✓

As the data came back from National Centers, so the International Committees met again and prepared for the next round of data collection/pre-testing. After pre-testing they met again for the finalization of the instruments. Committees met on an average of two or three times a year. Between meetings, however, committee members carried out a great deal of work on their own.

At the same time work was undertaken continually by the IEA Central Staff together with some consultants on the construction of questionnaires, attitudinal and descriptive scales. (See Chapter 8).

In order to provide a more detailed and more robust conceptual framework for the IEA and similar studies, two meetings were held under the title of "The Cross-National Conference on Education, Manpower and the Economy" (U.S. Office of Education Grant No. OE-1-061967-1989). In these meetings outstanding scholars from the various social sciences were brought together to review the type of research being undertaken, and at the same time to suggest from their own discipline point of view the types of hypotheses which could be decided in such surveys, together with the types of variables which had been missing from the Mathematics Study, but which could be included in further studies. These were fed to those people responsible for the questionnaire construction, (See Chapter 8).

The Central Staff and the Consultants also worked out the whole procedures for the administration of the field testing, and these have been written up in Manuals 1, 2 and 3 (See Volume II, Booklet I).

At various junctures throughout Stage 1, Bulletins were issued, which brought together the main technical papers and policy decisions taken over a period of months of most importance to National Centers, such that all of these were within one document. From the beginning of 1968 newsletters were sent out by the Chairman every quarter. Two examples of these are given in Appendices B-1, B-2.

The content analysis in most subject areas was finished by the end of 1967, and the pre-tests produced by Spring 1968. The pre-testing took place between February and October 1968 and the item analyses were returned between June and December 1968. Final versions of the instruments were ready between December 1968 and March 1969.

1.6 IEA Organization

In order to be able to receive money and to become a legal organization in its own right, the Council of IEA decided that the organization should become incorporated. It was decided to become incorporated in Belgium, following advice from the UNESCO Legal Department. IEA thus became formally incorporated on June 13, 1967, and the Statutes are attached. (See Appendix C). The structure of IEA is given in detail in these Statutes. The Council has met during Stage 1 approximately once a year, as has also the Standing Committee. The Bureau, consisting of Professor Torsten Husen, Professor B.S. Bloom, and Mr. D.A. Pidgeon, has met approximately every two to three months to take the major decisions required between Council meetings.

Dr. Postlethwaite has been responsible for the day-to-day administration and for heading up the permanent staff. On April 1st, 1968, Dr. B.H. Choppin of Cornell University was employed as full-time data processor, and on January 1st, 1969, Mr. J.K. Hall, Head of the Surveys Department of the National Foundation for Educational Research in England and Wales also joined the IEA full-time staff to work in data processing.

1.7 Participation in IEA

Booklet 6 of Vol. II is the IEA brochure which gives full details of all persons involved both internationally and nationally in IEA.

Canada, Hungary, Israel, the Phillipines, and New Zealand have also applied to IEA. The U.S.S.R. was formally accepted into IEA in December 1967, but has not found it possible actually to participate in any of the work so far and negotiations are still continuing with other countries.

In December 1968 National Centers committed themselves definitely to Stage 2, Science, Reading Comprehension and Literature as given in Table 2 as well as tentatively to English as a Foreign Language, French as a Foreign Language and Civic Education for Stage 3.

TABLE 2

	STAGE 2													STAGE 3													
	SCIENCE					READING COMPREHENSION				LITERATURE				FRENCH					ENGLISH				CIVIC EDUCATION				
	I	II	III	IV	IVS	I	II	III	IV	II	III	IV	IVS	I	II	III	IV	IVS	II	III	IV	IVS	I	II	III	IV	IVS
Australia	x	x		x	x																						
Belgium		x		x		x	x		x	x		x															
Chile	x	x		x	x	x	x		x	x		x					x				x						
England	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x					x	x	x	x	
Fed.Rep. Germany	x	x		x	x														x		x		x	x			x
Finland	x	x	x	x		x	x	x	x	x	x	x							x				x	x			
France				x																							
Hungary	x	x		x		x	x		x																		
India	x	x		x		x	x		x																		
Iran	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x		x	x	x		x	x	x	x	
Italy	x	x		x		x	x		x	x		x	x		x		x		x		x		x	x		x	
Japan	x	x		x	x																						
Nether-lands	x	x		x		x	x		x						x		x		x		x						
New Zealand		x		x	x		x		x	x		x	x														
Poland	x	x		x		x	x		x																		
Scotland	x	x		x	x	x	x		x						x		x										
Sweden	x	x	x	x	x	x	x	x	x	x	x	x				x	x		x	x	x		x	x	x	x	x
Thailand	x	x	x	x		x	x	x	x										x	x	x						
U.S.A.	x	x		x		x	x		x	x		x	x	x	x		x	x					x	x		x	x

CHAPTER 2

SCIENCE

2.0 Introduction

Several factors influenced the pattern of the work in Science.

(a) Science is now almost universally recognised as an essential component in the education of pupils in secondary schools and, at an appropriate level, is being increasingly introduced into the curricula of primary schools.

(b) In most countries the content and method of science education are undergoing critical examination and appraisal; syllabuses are being changed and new emphases are being given to the methods of study. In brief, one can say that school science is being given a more contemporary outlook, and this means that much new, and often untried, material is being introduced and much traditional material discarded, also that a spirit of scientific enquiry is being infused into the work, in which methods of working and the attitudes developed are now regarded as at least as important as the knowledge acquired.

(c) The traditional differences between countries, and between schools within countries, are widening because of the vastly different stages in reform that have been reached. Thus in some countries, where the curriculum is in the hands of the individual school or the Local School Board, some schools will have had several years of experience of the new materials and methods while others will still be following the traditional science courses.

(d) Probably the most significant difference between the science education of one country and another lies in the nature and extent of the practical work undertaken in the laboratory. At one extreme laboratory experiments are regarded as the essential experience upon which the theoretical structure of the subject is erected. At the other, practical work is seldom, or never, attempted and when it is, its purpose is to illustrate or reinforce what has already been learned.

It is clear that science education is in a state of flux, healthy as regards the pupils being educated but imposing serious difficulties from the point of view of an objective, comparative study. Obviously the first thing to be done was to obtain, as far as was possible in the circumstances, a synoptic view of the science curricula of the participating countries, both present and what was likely in the immediate future.

2.1 First Stage

Accordingly a form (Appendix D-1) setting out a list of subject areas and teaching objectives was devised for completion by the National Centers for each target population. In this way it was hoped to obtain sufficient areas of agreement between countries on which to base cognitive tests measuring the attainment of specific teaching objectives.

Several points need to be made in connection with this form.

(a) The number, nature and method of expression of the subject areas.

An important decision was taken at the very beginning, viz., that with the exception of the so-called science specialists of Population IV S, all pupils would take a test in Science comprising unidentified sections of earth sciences, biology, chemistry and physics. In other words, science education was to be seen as a whole, even though this meant that some pupils would be unable to answer certain items in their papers.

The form in which the subject areas were set out implied a certain approach to the teaching which might not have been equally appropriate to every country and certainly differed from one branch of science to another. Similarly the listed subject areas were not comparable in extent or importance. Thus, whereas in physics the subject divides itself fairly naturally under such headings as heat, mechanics and electricity, the arrangement of the biological subject areas under three levels of organisation, the cell, the individual and the population, is of a different kind. Nevertheless, the list of subject areas was agreed upon and subsequently functioned satisfactorily.

(b) The teaching objectives

Two totally different approaches were possible. On the one hand a list of objectives as they might be set out by a practising teacher concerned only with his actual job of teaching, and, on the other, a list arranged as a hierarchical structure which would facilitate subsequent analysis of test items. The first alternative was adopted, but comments from National Centers made it quite clear that in the long term this was a mistake, although from the point of view of gaining an impression of the science education in the different countries the objectives as set out were not without merit.

(c) The present and future positions

National Centers were asked to assess the importance attached to each topic on a three point scale and to attempt to identify the teaching objectives related to it. It was recognised that in many cases the replies would be largely impressionistic and subjective but at this point it was essential to collect as much information as possible.

The task of compiling the subject area/teaching objective grid was a formidable one undertaken at its first meeting by the International Committee for Science; the grid was amplified by two tables (Appendix D-2 Table I and Table II) which expanded the list of objectives and the list of subject areas as might be appropriate for the three target populations respectively.

At the same time a preliminary list of the areas of interest and the hypotheses to be tested was drawn up and circulated to National Centers so that they could gain from the beginning an overall view of how the science side of the project was being developed. They were set out as follows:

I

1. To what extent are the newer science teaching methods based upon investigation and pupil enquiry achieving better results than traditional methods?

2. Are any differences observed more marked in achieving some objectives than others?

3. To what extent is the success of science teaching, and particularly the newer methods of science teaching, dependent upon

   a. the material conditions under which science is taught, e.g., provision of laboratories and apparatus,

   b. the time allocated to science and the fraction of it spent in practical work in the school timetable,

   c. the provision of laboratory assistance,

   d. the training teachers have received and the facilities for refresher courses?

4. Is a more adventurous attitude towards future careers in science, as indicated by a wider range of desired careers, favored by the newer teaching methods?

II 1. What differences can be detected between the sexes in their attitudes to, and their achievements in:

   a. science as a whole,
   b. the physical sciences and the life sciences.

2. To what extent are any of these differences related to:

   a. social variables,
   b. home background, including conditions and interests,
   c. school administration and organization,
   d. teacher variables?

III 1. To what extent is achievement in, and attitudes towards, science related to:

   a. the level and rate of advance of technological development in the countries concerned,
   b. rural and urban environments?

IV To what extent is achievement in, and attitudes towards, science related to school organization variables such as comprehensive and selective systems of education?

2.2 Second Stage - Topic and Objective Specifications

Ten National Centers completed the grids by the specified dates and in addition they, and one or two that were not able to complete the grids in time, contributed valuable comments on both subject areas and teaching objectives. As a result, a shortened list of subject areas reasonably common to all those countries from whom information was available was drawn up for each population and sent to each National Center (see Appendix D-3) with an urgent request that test items covering the specified common subject areas and judged to be suitable for the population specified in the country of origin should be submitted without delay.

At this point the weakness in the expressed objectives from the point of view of test construction was felt acutely and National Centers were given some guidance on procedure and asked particularly to produce items measuring the higher scientific abilities and those testing special abilities such as the design of experiments or the handling of scientific apparatus.

About 2000 items from twelve different countries were received in time for the second meeting of the Science Committee when the first draft tests were to be compiled.

Most of the items received were concerned, as requested, with the common subject areas and therefore potentially usable, but they were very uneven in quality and a high proportion were not in a suitably objective form in which they could be marked mechanically. More serious still, however, were the facts that far too few items measuring higher abilities were received and that there were very large discrepancies, both in quantity and quality, between those contributed by different countries. It was difficult, therefore, to achieve the desired balance in the construction of the draft tests and to give a fair representation to every country. There was, however, the safeguard that all the items were to be pre-tested in each country and that National Centers would have ample opportunity for further comment.

Modification of the scheme of teaching objectives

As indicated above, the response from National Centers and the need to categorise test items had suggested that it was essential to devise a scheme of objectives arranged in some hierarchical pattern and adapted to the particular nature of science education. Alternative schemes had been submitted by two National Centers and these were considered in relation to the well-known scheme given in "Taxonomy of Educational Objectives." As a result a scheme of four hierarchically arranged objectives was adopted. These were stated briefly as A. Functional Information, B. Comprehension, C. Application, D. Higher Processes (see Appendix D-3for amplification).

Two points about this scheme deserve comment in addition to that given in the appendix. Category A, Functional Information, was intended to eliminate items of pure recall as yielding little of value to the project, and, at the same time, to make room for certain practical abilities needed for the acquisition of scientific information, such as observation and the interpretation of observations. Category D, Higher Processes, was introduced to avoid the difficulty that, in science, the categories Analysis, Synthesis and Evaluation of the Taxonomy cannot be arranged as a simple hierarchical system.

It is intended that the categorisation of test items selected for inclusion in the final tests should be validated in a separate operation carried out by specially briefed teachers in each country. In the meantime the tests were built up on the basis of objective categories agreed by the Science Committee.

2.3 Third Stage - Pre-testing

Selection of test items for comment and eventual pre-testing

At a meeting of the Science Committee held in Hamburg in September 1967 a first selection from the items submitted by National

Centers was made. First the items were selected from the point of view of appropriate subject area coverage, a pre-determined balance of teaching objectives and, as far as possible, equal representation of the contributing countries. The final criterion applied was that the item was, in the view of the Committee, potentially a good one. The items were then put into a reasonably common form (multiple choice with five alternative responses) and new items devised to fill the most obvious gaps. It must be stated, however, that item construction is a skilled process and that because of the limited amount of time available to them the Committee was forced to accept some items which had obvious deficiencies.

Following this meeting, very rough drafts of the pre-test forms were sent to National Centers for comment. In compiling these forms the balance between the items assessing the four behavioral categories was aimed roughly in the proportions set out below, but in fact this balance was not achieved.

Population I	A3	B2	C1	D1
Population II	A3	B3	C2	D1
Population IV and IVS	A1	B1	C1	D1

The pre-testing

In spite of a very tight time schedule, helpful comments on the pre-test drafts were received from almost every one of the participating countries and, on the strength of these, a small group of three members of the Science Committee prepared the final pre-test forms. The items were now much improved in quality, their presentation was more uniform and, in general, they gave the impression that real progress was being made towards a valid, final testing instrument.

The following table summarises the materials sent out for pre-testing:

Population	No. of Sub-Tests	Number of Items: Total	Behavioral Category				Major Subject Area			
			A	B	C	D	Earth Science	Biology	Chemistry	Physics
I	4	104	38	35	18	13	13	40	(51)	
II	8	192	54	66	35	37	1	65	64	62
IV	8	126	37	22	36	31	19	46	30	31
IVS Bio.	4	96	20	28	18	30		96		
IVS Chem	4	92	19	41	19	13			92	
IVS Phys	4	92	16	35	21	20				92

Pre-testing of science items was carried out by sixteen countries early in 1968 on the basis of the scheme set out below:

		Population I	Population II	Population IV
Australia, Belgium, Germany, India	Sub-tests	1 2 3 4	1 2 3 4	1 2 3 4
Chile, England, Iran, Netherlands	Sub-tests	1 2 3 4	4 5 6 7	4 5 6 7
Finland, Italy, Japan, Scotland	Sub-tests	1 2 3 4	1 2 7 8	1 2 7 8
Poland, Sweden, Thailand, U.S.A.	Sub-tests	1 2 3 4	3 5 6 8	3 5 6 8

For the specialist tests in Science (4 Biology Sub-tests B1, 2, 3 and 4; 4 Chemistry Sub-tests C1, 2,3 and 4; 4 Physics Sub-tests P1, 2, 3 and 4), a slightly different rotation was required.

SPECIALIST TESTS

Australia, Belgium, Germany	B1	C1	P1	B2	C2	P2
India, Chile, England	B1	C1	P1	B3	C3	P3
Iran, Netherlands, Finland	B1	C1	P1	B4	C4	P4
Italy, Japan	B2	C2	P2	B3	C3	P3
Scotland, Poland	B2	C2	P2	B4	C4	P4
Sweden, Thailand, U.S.A.	B3	C3	P3	B4	C4	P4

Before undertaking their pre-testing programs National Centers were given advice on how to deal with difficulties of translation, the question of the use of popular and scientific units and of unfamiliar plants, animals and materials. In general, National Centers were given a reasonably free hand to substitute, at their discretion, suitable alternatives within the structure of the item, on condition that the Co-ordinator was informed and that copies of the translated items, as set, were made available to the Science Committee for their use in drawing up the final tests. Singularly few difficulties occurred; those that did were fairly easily overcome and most of them arose from the correct alternatives in the question being given away in the process of translation. It might be amusing to speculate on the reasons for the absence of difficulties due to purely local conditions. Why should the science taught be much the same all over the world?

National Centers also received guidance on the administration of these cognitive tests so that certain standard procedures were followed and the same kind of statistical data were produced.

Pre-testing: the item analysis

The results of the pre-testing, carried out on judgment samples of 100-200 pupils for each population and sub-test, were analysed independently by National Centers and submitted to IEA Headquarters in Hamburg on a prescribed form. In addition, the results for each item from different countries were collated on a single form.

The Science Committee had thus available to them something in the region of 700 items, for each of which a difficulty value, and a discrimination index for each country that had been able to carry out the pre-testing to time were known. The range of difficulty for each item between countries was also known. The number of countries which was able to submit their analysed data varied from one sub-test to another, but in most cases it amounted to at least six and the combined results can be regarded as providing a satisfactory basis for the final selection of the test items.

Selection and final editing of the pre-tested items

There were roughly four times as many pre-tested items as were needed although these were unevenly distributed between the four behavioral categories and some had been shown to be unsuitable on statistical grounds. A fair degree of selection was therefore possible and there would still be a reserve of items that could be used on some future occasion.

The final selection of items and their arrangement in tests was carried out by the Science Committee at a meeting in Hamburg 1st-10th July 1968. Their aim was to select items in such a way as to achieve a spread of difficulty which meant that each country was fairly represented and to use only those items which discriminated well. At the same time the subject areas and the behavioral categories had to be suitably represented.

For each population, therefore, a frame was constructed in which the rows represented subject areas and the columns the behavioral categories or teaching objectives. The aim was to have at least six items in each sub-test that could be drawn from the frame for the purposes of separate analysis. For Populations I, II and IV the subject areas chosen for this purpose were earth sciences, biology, chemistry and physics, although within this very broad classification adequate representation was given to practically all the topics listed as being reasonably common to all countries as a result of the original analysis of the curricula grids. For Population IVS sub-division of the separate sciences was of course possible.

Where other factors were equal, items were selected which were judged to give information about the scientific climate in which a student was being educated in preference to those that measured the effect of specific instruction.

2.4 Construction of the final tests

Several other considerations had to be borne in mind in the compilation of the final drafts.

Anchor items

The first was the need to include anchor, or bridge, items which were common to two populations and which would therefore allow comparisons between populations in point of time to be made. In the final tests there are 11 such items between Populations I and II and 11 between Populations II and IV.

Rotating sub-tests of selected Population IVS items

In order to meet the difficulty of defining a science specialist in the pre-university years, for whom the tests for Population IVS have been devised, and to enable comparisons to be made with corresponding students in countries where such students are not easily identified and where the Population IVS tests will not be taken, some items common to Populations IV and IVS are desirable. So that an adequate field of subject matter can be tested within the time allocated and without affecting unduly the nature of the tests for Population IV, a system of rotating sub-tests has been introduced. Six sub-tests each of two biology, two chemistry and two physics items have been added in rotation to the 24 items in the second test of Population IV. Each student in this population will therefore respond to six IVS items, but for each group of six students, 36 IVS items, twelve biology, twelve chemistry and twelve physics, will be administered altogether.

Tests on understanding the nature of science

Representations were made by one or two countries that an attempt should be made to assess the students' ability to understand the nature and methods of science as distinct from the purely cognitive aspects of science. With this end in view, a test (SST 1 - Supplementary Science Test 1), which drew very heavily on the TOUS tests devised at the University of Chicago, was compiled and pre-tested in September/October 1968.

Comments were received from eleven countries and full pre-test results, including item analysis, from eight. On the strength of this information it was decided to include a separate test on "Understanding the Nature of Science" for completion with the questionnaires on the affective measures by Populations II and IV. The tests for both Populations will consist of 15 items, for which 20 minutes of testing time will be allowed. Seven items will be common to both Populations.

Tests on practical abilities in science

It was stated earlier that major differences occurred between countries with regard to the place accorded to practical work in the laboratory or field. Many of the new developments in science education are concerned with the question of the nature and extent of the first-hand experience that is desirable during the study of science at school. One of the most important hypotheses to be tested by the science project is in fact, that science teaching methods based upon actual enquiry by thoroughly sound scientific methods achieve better results, even measured in cognitive terms, than traditional methods. It had been hoped that suitable items for measuring these abilities would have been found amongst those submitted by the various countries and a strong appeal appeal was made for them by the Science Committee. Since none were forthcoming, a grant was made jointly by IEA and the NFER England for a study of the problem, and the production of suitable items. This study was based at the University of Leicester, England. As a result two tests SST 2 and SST 3 (Supplementary Tests 2 and 3) were devised and pre-tested together with SST 1 in September/October 1968. Some of the items are of a paper and pencil type; others require a minimum of apparatus.

Results of the pre-testing of these "practical" tests were received in time from three countries only, although comments were received from several others. It seems clear that some countries do not attach as much importance to practical work as others, while the administration of tests requiring even a minimum of equipment and laboratory space presents difficulties that few countries expressed willingness to undertake. In consequence, it was decided to incorporate ten "pencil and paper" items aimed at measuring the results of practical experiences in each of the two cognitive tests for Population II and twelve similar items in test IVA for Population IV. Eight of these items will be common to both Populations.

In addition, five "practical" laboratory tests involving only a small amount of very simple equipment will be offered as a national option. These are all planned for one and a half hours in the laboratory although it is expected that the actual manipulations and recordings will be completed within an hour. The intention is to remove from the student all sense of undue pressure. One of these five tests will be in science, equally divided between biology, chemistry and physics, for Population II. The other four, in science, biology, chemistry and physics respectively, will be for Population IV as school options within the national system.

A summary of the final science tests is given on the next page.

A Summary of the Final Science Tests

Population	Name of Test	No. of Items	Time Minutes	Comments
I	Science I A	20	30	Most of the knowledge and understanding tested is not likely to be the result of specific science teaching but of the awareness of the opportunities provided by the environment in general elementary teaching.
	I B	20	30	
II	Science II A	40	60	Both tests contain "paper and pencil practical" items.
	II B	40	60	
IV	Science IV A	36	60	IV A will contain "paper and pencil practical" items.
	IV B	24 +6 rotating IVS items	60	Each set of six rotating items will contain two items each on biology, chemistry and physics and will be answered by one student in every group of six students.
IVS	Science IVS Biology	40	60	A student in a country which has elected to take these tests will take either one or two of them at either his own choice or that of the school.
	Science IVS Chemistry	40	60	
	Science IVS Physics	40	60	

It is suggested that countries intending to test Population III should use tests II A and II B.

As a national option the following "practical" tests are offered:

Population	Name of Test	No. of items	Time Minutes	Comments
II	Science II Practical	9 (score 15)	90	
IV	Science IV Practical	7 (score 19)	90	
	Science IV Biology	14 (score 14)	90	The raw scores to be adjusted to a total of 40 and rounded off.
	Science IV Chemistry	3 (score 18)	90	
	Science IV Physics	7 (score 12)	90	

The distribution of items by behavior, content, and populations are shown in the following table.

Population	No. of Sub-Tests	Number of Items											
		Total	Behavioral Category							Major Subject Area			
			A	B	C	D	I	II	III	Earth Science	Biology	Chemistry	Physics
I	2	40	12	12	7	9				9	13	4	14
II	2	80	19	20	7	14	10	1	9		19	19	22
IV General	2	66	15	12	12	9	3	1	8		16	16	16
National Options													
IVS Bio	1	40	6	10	12	12					40		
IVS Chem	1	40	8	16	8	8						40	
IVS Phys	1	40	8	10	10	12							40
IV Bio Practical	1	15					3	10	2				
IV Chem Practical	1	18						18					
IV Phys Practical	1	9					2	1	6				
II Gen Practical	1	17					6	8	3				
IV Gen Practical	1	16					0	13	3				

2.5 Student, Teacher and School Questionnaire

Concurrently with the production of the cognitive measures questionnaires were developed which will provide some of the information required when the hypotheses are being tested. Those relating specifically to science, ST 1 and 2 - Science and Teacher Questionnaire - Science are shown in Volume II, Booklet 4.

2.6 General Appraisal

The science part of the project has reached the present stage as a result of the very great efforts made by the National Centers and the Science Committee backed by the very efficient Secretariat, and partly by the calculated risk taken by the Science Committee in putting out for comment and trial materials with which they were by no means satisfied, but which they felt to be potentially satisfactory. Criticism was often harsh, occasionally unhelpfully destructive, but sometimes constructive and valuable. In spite of the initial difficulties an enormous amount of work has gone into the production of these instruments and there is now a general feeling that the instruments in science are very adequate for the purposes for which they have been produced.

CHAPTER 3

Reading Comprehension

3.0 Introduction

The belief that it might be reasonable to include reading comprehension as one aspect of further cross-national surveys of educational achievement stemmed from the initial pilot international study (Foshay et al., 1962). In this exploratory study, a reading comprehension test was included as one element in the test battery. It was then recognized that translation into eight different languages might give rise to severe problems, but the task was undertaken. As the results were examined, it was found that between-country variance in mean score was relatively small, and furthermore the relative difficulty of single items remained quite stable from one language to another. The average correlation of item difficulty indices from language to language was approximately .90. An item that was hard in one language tended to be hard in others.

Further work at the end of secondary school level (Kumbaraci, 1966), comparing difficulty in English and Turkish, tended to support the earlier findings. However, in this older and more select group, the correlations were about .70 rather than .90. Also, comparison of items, with respect to their ability to discriminate high-ability from low-ability students, indicated that this aspect of the items was quite unstable from the one language to the other.

All in all, the results on using reading tests as translated seemed sufficiently promising so that further work in this area was proposed. The further work would use the larger and more carefully selected samples that were characteristic of the IEA mathematics study (Husen ed., 1967), and would gather extensive data on individual background, school characteristics, and national educational programs in order to study relationships between these factors and reading achievement. Therefore, a Reading Comprehension Committee was set up within IEA.

Original members of the Committee were:

- Prof. Robert L. Thorndike, Teachers College, Columbia University, U.S.A., Chairman.
- Dr. Lily Ayman, Teachers College, Tehran, Iran.
- Mlle. Francoise Bacher, Institut National de l'Orientation Professionnelle, Paris, France.
- Mr. Alan Brimer, Director, Educational Research Institute, Institute of Education, University of Bristol, England.
- Mr. Alan Murphy, Institute of Education, University of Ibadan, Nigeria.

Mr. Murphy found himself unable to continue after the initial meeting of the Committee, and his place was taken by Dr. Jan Vastenhouw, Institute of Education, University of Amsterdam, The Netherlands.

3.1 Preparation of Rationale

Initial planning for the study of accomplishment in reading comprehension in different countries was carried out at a meeting of the IEA Reading Comprehension Committee in Hamburg in November, 1966. At that time, a statement of rationale for and procedures tentatively to be carried out in a reading comprehension study were developed. This rationale was revised at a June 1967 meeting of the Committee on the basis of comments received from different national groups and, in its revised form, is attached to this report as Appendix E.

As indicated in Appendix E , several considerations led the Committee to focus on the cognitive aspects of reading rather than the aesthetic and affective, and upon expository rather than narrative material. Practical realities restricted the assessment to paper-and-pencil group testing, and this, together with the focus of interest of those working in other subject areas, ruled out testing of children before the age of 10 years.

The above constraints ruled out much of what would have been most interesting in the field of reading, to wit, the early development of reading skills. The focus was upon comprehension by relatively mature readers, rather than upon the initial stages in development of decoding skills.

It was initially proposed that we test both ability to get information from a text when the text is present to be referred to and ability to answer by recall questions based on a text that had previously been studied. The second type of testing was abandoned in the final instruments, primarily because of lack of time--both lack of time to prepare and try out the test materials and lack of testing time upon the occasion of the final test.

Word knowledge was seen as both a condition of and an aspect of reading skill, so a word-knowledge test was developed as described more fully below. A reading speed test was also developed, since speed seemed a relevant second dimension of reading competence.

The statement of rationale was sent out to the National Centers to elicit comments from the national Reading Comprehension Committees. In addition to being asked to comment on the general plan and rationale for the work in reading comprehension, each National Committee was invited to submit reading passages, especially ones with test items already available, that might feed into the final test of comprehension as the test was developed. These materials were intensively reviewed at a meeting of the Reading Comprehension Committee in Paris in June, 1967. Passages were rated as to their apparent difficulty by members of the Committee, and sorted into seven levels that were described as 8-year-old, 10-year-old, 12-year-old, 14-year-old, 16-year-old, 18-year-old, and 18+. The designations represented the age level at which it was judged the passages would be appropriate for the typical student. However, the judgments were intended to represent only a rough preliminary grading of the passages.

3.2 Preparation of Materials

A number of passages that seemed to the Committee members to be suitable in style and content, and not to be peculiar to any one culture or country, were selected as possible materials for the test. These passages were reproduced and circulated to all of the countries for comment, and only those passages that were almost universally considered to be acceptable for inclusion were retained for further development. During the summer of 1967, an extensive assortment of new test items was prepared, primarily by the chairman of the Reading Comprehension Committee, to supplement those that were already available for the passages. (Many existing items were modified and edited to cast them into a multiple-choice format that would permit machine scoring.) These items were subsequently sent to all National Committees for comments and suggestions and for the supplying of additional items wherever a National Committee cared to make such a contribution to the enterprise.

Taking account of the comments and suggestions of National Committees, and including items provided by these groups wherever possible, six try-out test forms were prepared for each of the populations to be included in the IEA study to wit, 10-year-olds, 14-year-olds, and students in the final year of secondary education. Each 10-year-old try-out test included three passages, one that had been rated as an 8-year-old passage, one a 10-year-old passage, and one a 12-year-old passage. The try-out tests for Population II (14-year-olds) contained four passages spanning the 12-, 14-, and 16-year-old levels. The end-of-secondary test also included four passages, but at the 16, 18, and 18+ levels. Thus, the test at each level had one passage in common with the next level. All told, 50 different passages and 445 different items were tried out.

3.3 Pre-Testing

Tests were distributed for try-out approximately in March, 1968, and try-outs were completed before the end of the 1967-1968 school year. The amount of material to be translated and tried out was more than could have been handled in any one country, so the six forms of the tests were divided into three sets of two forms each, and each participating country was asked to pre-test one-third of the material. The schedule called for each test to be tried out in four countries, one an English-speaking country and the other three each representing a different language. It was hoped that in this way each test would have been pre-tested in enough different languages to make it possible to eliminate items that were discriminating in only one or two languages, and that the items that were retained would prove to be discriminating in the remaining languages as well as in those in which they were tried out. For most tests, data were available as planned for four countries, but difficulties arose in relation to some of the try-outs so that a few passages were tried out in only two or three countries.

3.4 Preparation of Final Tests

Item statistics on the try-out data were returned to IEA headquarters at Hamburg, where they were collated and made available to the chairman of the Reading Comprehension Committee who in turn organized them in relation to specific passages and distributed them to the members of the Reading Comprehension Committee. Committee members were asked to indicate which passages should certainly be kept, and which should be the ones to be eliminated. Within passages that might be kept they were invited to nominate items for retention and for rejection. Final selection of the passages and items became the responsibility of the Committee chairman.

Passages were selected to provide a suitable range of difficulty for the population in question. A second consideration in selection of passages was the number and diversity of items showing suitable difficulty and discrimination. Finally, the attempt was made to get variety of content and treatment. Items were retained that showed acceptable difficulty and discrimination indices, and within these limits an attempt was made to include items covering as wide a range of reading skills as possible, i.e., items dealing with specific factual details, with the main idea, with inference beyond the literal content of the passage, with the author's point of view and purpose, and with the author's writing techniques.

Final tests were prepared as follows: For Population I (10-year-olds) two tests each planned to have a time allowance of 25 minutes, covering 8 passages and including a total of 45 items; for Population II (14-year-olds) two tests each planned to have a time allowance of 45 minutes, covering 8 passages and a total of 52 items; for Population IV (end of secondary school) two tests each planned to have a time allowance of 50 minutes, covering 8 passages and a total of 54 items.

3.5 Reading Speed

The original rationale for the reading appraisal included a proposal to assess reading speed as well as reading comprehension, since speed is a relatively independent aspect of reading skill that has some importance as an academic accomplishment. Assessment of reading speed has always presented a number of technical problems and no fully satisfactory technique for this type of assessment has been developed. The plan originally proposed by the Reading Comprehension Committee was to appraise reading speed by presenting passages into which words had occasionally been inserted that spoiled the sense of the passage and had no relationship to what went before or after. The examinee was to be instructed to underline or cross out these words as he came to them, and the number of such cross-outs in a closely-timed test would indicate the speed at which he was reading.

One problems with this technique is that it is a somewhat unnatural one, implying a set to proofread rather than to read for sense, and another difficulty is that if the material is of a fairly high level, the speed may be very much slowed down by comprehension difficulties. Somewhat unilaterally, the chairman of the Reading Comprehension Committee modified this original pattern to one that has been widely used with young children in tests of reading speed in the United States. This procedure involves providing a series of relatively short paragraphs, each of which ends with a question and is followed by three words, one of which is to be underlined as answering the question. The paragraphs have to be constructed in such a way that the most economical and efficient way of answering is to read the whole paragraph, rather than reading merely the question. When this is the case, the process is a reasonably natural one of reading a sentence or two followed by a question and then responding to that question.

In the materials that were prepared for this purpose in the IEA study, the sequence of paragraphs constituted a single continuous story. The attempt was made to write the material at a very low level of difficulty, so that the measure would be that of the fluency in mechanics of reading material of a very simple nature. Two stories were prepared and were distributed to the national groups for comment and subsequently for try-out. The materials were administered with the instruction for each child to read as rapidly as he could, while still getting the information needed to answer the questions. The child was instructed to mark clearly where he was at the end of three minutes and again at the end of six minutes so that it would be possible to get two distributions of items attempted and provide two marks to guide the final decision as to the optimal testing time.

In addition to getting data on the timing of the passages, each country was asked to tally the number of errors on each of the items in the stories so that items that were ambiguous or too difficult could be identified and modified. It turned out that one story had a number of rather ambiguous items and items that presented difficulty to children whereas the other had relatively very few. On this basis, the second story was chosen for the final testing and was edited to try to minimize those difficulties that were in fact found in it. The first story was also edited to eliminate as much as possible the ambiguities and difficulties, with the intention that it would be used as a practice story. The time limit that seemed appropriate for final testing with the speed test was four (4) minutes.

Certain questions arose in discussing the processing of the data from the speed test. These have to do with a practical method of scoring, since the test did not use a separate answer sheet and scoring for correctness would be quite laborious. It was initially intended that the test would be scored merely in terms of the last item attempted, since it was intended that there would be a very low error rate on the test. In practice, it turned out that in a few of the countries the story was a good deal more difficult than had been expected and the error rate was fairly substantial. The problem of how to adjust a speed score for errors when these are fairly frequent is one that is not easily solved. Conceivably, one could score merely the number of correct responses, or one could even introduce a penalty for errors so that the score for effective reading speed might be the number right, minus some fraction of the number wrong. Any of these scoring procedures introduces very serious practical scoring difficulties, and this problem remains not fully resolved at the present time. What seems a practical compromise is to count errors on only the first page of the test, thus getting an error score, but quite an unreliable one for each individual. The speed score would be the last item attempted. Having both scores available, one could study them separately or in any weighted combination. Hand-scoring only the first page for errors should go very quickly and not prove burdensome for National Centers.

3.6 Word Knowledge Test

A third component that was proposed in the original plan for the reading test was a test of word knowledge. This was seen less as an inherent part of the reading test than as a control variable that might be useful and relevant in relation to all the tests of the IEA study.

As something of a tour de force, the Reading Comprehension Committee undertook to try to produce a common test to be used across languages in all of the different participating countries.

The item format that was selected was that of word pairs where the words of a pair must be judged to be approximately either synonyms or antonyms, that is, either nearly the same in meaning or nearly the opposite in meaning. It was thought that this item format would be easier to translate into different languages than the conventional multiple-choice format with a number of wrong options. Here it was felt that the attractiveness of wrong options would vary so much from country to country that it would be extremely difficult to get equivalent items from one language to another.

In order to accumulate a pool of items that might be used for this study, National Committees in each country were invited to submit sets of word pairs ranging from easy to difficult, providing them to the Reading Committee in their English translation. A list of some 300 word pairs was assembled. The difficulty of these in English was judged by several individuals at Teachers College, and the words were arranged into 10 classes with respect to their judged difficulty. The set of words was then circulated to each of the National Committees, and each Committee was asked to indicate which of the pairs would be either impossible or difficult to translate into a pair of corresponding difficulty in the language of the country.

The data from the above assessments were assembled and tabulated, and all word pairs were dropped from further study in which more than one country indicated difficulty in the matter of translation. The remaining word pairs were assembled into three try-out tests, one for 10-year-olds, one for 14-year-olds, and one for end of secondary school and were circulated to the participating countries for try-out in their school systems. Some of the items in adjacent tests were common to permit scaling items from successive test levels on a common scale.

The original hope and plan in designing the vocabulary test was that there would be among the rather large pool of items tried out a sub-set of items that were of comparable difficulty from one country to another. The problem becomes one of defining comparable difficulty in different countries using different types of samples and different languages. The assumption on which the attempt to scale difficulty across countries was based was that on the average for the whole pool of items they would be about equally difficult in English and in the language into which they were being translated. The subsidiary expectation was that the correlation of item difficulty from one language to another would be positive and fairly high, though obviously far from perfect.

Data were reported to IEA in the form of percent passing a given item. In some countries, item vs. test correlations were also reported, making it possible to identify items that were not differentiating between the more and the less able pupils.

Items were scaled for difficulty, first within each age sample, then across age samples for a country, and then in relation to the U.S.A. as a common anchor population. The sequence of steps is outlined below:

(1) Proportion "knowing" an item was defined as proportion correct minus proportion wrong (R-W).

(2) The proportion established in (1) was converted into a deviation score on the baseline of a normal curve.

(3) Using those words that were common to the Population I and Population II tests, the difference in means and ratio of standard deviations in the two groups was estimated.

(4) Scale scores for the Population II items were converted to the Population I scale by the appropriate linear equation.

(5) In the same manner, Population IV scale values were converted to Population II and finally to Population I scale values.

(6) Scale values as determined in (5) for Country X were plotted against U.S.A. scale values, and the means, standard deviations, and intercorrelation computed.

(7) A conversion equation was applied to the Country X data to produce the same mean and standard deviation as those for the U.S.A.

As the data were received and examined, it seemed rather doubtful as to whether either of these original expectations were borne out. In particular, there were a number of the translation languages in which the items showed a very limited spread of difficulty. Generally speaking, the range of difficulty was greater in English than it was in the languages into which the words were translated and this kind of a regression towards mediocrity was larger in some languages than in others.

It also appeared that the items tended to get systematically easier in translation, though this cannot be tested in any rigorous sense because one doesn't know much about the comparability of the samples from the different countries. It certainly did appear that the items which had been very difficult in English lapsed into being only moderately difficult in many of the translations, and that there were relatively few instances of the reverse, where an item that had been only moderately difficult in English became very difficult in one of the other languages. Furthermore, the average correlation of item difficulty from language to language, even including a range of words from those that were quite easy for 10-year-olds to those that were quite difficult for pupils at the end of secondary school (in English in any event), was only about .4 or .5. As a result, there was very large scattering of item difficulty, and it was hard to find a sub-set of items that were very nearly the same in difficulty across all languages. An attempt was made to select the best sub-set possible in the sense that items were chosen to try to balance difficulty from one language to another and items were chosen where the spread of difficulties was relatively small compared with others of the items. However, it must be admitted that the final result is one in which the equivalence from language to language is suspect at best.

Thus, to summarize, the following tests have been produced:

Population I	Reading Comprehension	Part I
		Part II
	Reading Speed	
	Word Knowledge	

Population II	Reading Comprehension	Part I
		Part II
	Reading Speed	
	Word Knowledge	
Population IV	Reading Comprehension	Part I
		Part II
	Word Knowledge	

The word knowledge test will be administered as part of the General Questionnaires in all countries.

The final instruments are to be found in Booklets 3 and 4 of Volume II.

CHAPTER 4

LITERATURE

4.0 Introduction

The IEA Study of achievement in literature has grown out of an effort begun in 1962 by Alan C. Purves, then at Barnard College, Columbia University, and A. W. Foshay, of Teachers College, Columbia University. These two colleagues considered whether the responses of students to literature could be analyzed cognitively; as a result Professor Purves developed a schema, "The Elements of Criticism and Interpretation" which sought to represent the field as a cognitive map (see Appendix F-1).

The schema was first subjected to critical scrutiny by a panel of well-known U.S. literary scholars and critics. Having survived this examination, the schema was further tested by using it to code brief essays written in response to an identical short story (translated as necessary) by secondary school students in Germany, Belgium, England and the United States. Once more, the schema survived; three independent readers coded the sentences in these essays with two-way agreements on 80% - 90% of the students' sentences, and three-way agreements on 60% - 70% of the sentences.

At this point, Professors Purves and Foshay proposed formally to the IEA Council that literature be included in Phase II of its work. The proposal was eventually accepted. An International Committee of five (see Vol. II, Booklet 6, p. 23) was appointed, and the IEA Council members from seven countries agreed to participate in this stage of the work. The countries were: Belgium, Chile, England, Iran, Italy, Sweden, and U.S.A.

In each of the participating countries a National Committee was appointed (see Vol. II, Booklet 6, pp.65-99), consisting of scholars and educational leaders in Literature in the mother tongue.

4.1 Preparation of Rationale

The IEA International Committee meeting on literature prepared a draft rationale for the study in 1966 (see Appendix F-2). This statement was circulated to the seven participating centers for criticism, and also to provide for them a framework within which they might carry out their first task: to prepare a statement of the goals of the teaching of literature in each country.

These statements of goals were returned to the International Committee, and became the basis for final decisions concerning the areas of achievement in literature to be measured.*

When the national statements of goals had been analyzed, it became apparent that there are, broadly stated, four categories of goals of instruction in literature:

1. Acquaintance with the principal literary works and the history of literature on one's own country and language.

2. The ability to respond skillfully and appropriately to literary works of several genres, (short stories, novels, lyric poems, epic poems, dramas, etc.).

3. The development of favorable attitudes towards serious literature, and a disposition to partake of it voluntarily.

4. The expression of a preference for literary works acknowledged to be of high quality.

4.2 The Selection of the Goals

Two of the four goals do not lend themselves to international testing. Goal number one - acquaintance with one's literary heritage, - is important, but unmanageable for international purposes. It is conceivable that some metadimension of achievement in this area could be identified and measured, but the development of such a construct seemed to the Committee so complex as to be beyond particability in the present project.

Goal number four - expression of a preference for literary works of high quality - could be measured with some difficulty, because preference refers both to content and style, and it is hard to hold one dimension constant. Moreover, on an international scale such a measurement runs the risk of being offensive. Such a test would have to venture into the highly personal world from which one's tastes arise. Moreover, the national committees and other advisors considered that such an attempt might be perceived as an effort to legislate taste, the last thing anyone wants. With some regret, therefore, the International Committee decided not to measure the achievement of this goal.

---

*Note: the US document reporting the goals of teaching literature was particularly thorough. It has achieved a considerable amount of attention among teachers of literature in the US. Requests for copies of this report have been received from several states, and it provided the basis for a chapter by Professor Purves on evaluation in literature, to be included in a forthcoming book, edited by Professor B. S. Bloom.

The remaining two goals - the ability to respond to literature, and the development of favorable attitudes toward serious literature - therefore became the basis on which the committee proceeded to the development of tests. The choice was defensible, because in each of the countries, these two goals were considered central. The first of them can be called "literary understanding," which in turn has both an achievement dimension and a preference dimension (understanding has slightly different definitions in different countries and in different critical schools within countries). The second can be called "literary attitudes." In order to develop instruments appropriate to the goals, the committee had to start from scratch. Almost no extended attention has been given to such measurement, even within nations. A cross-national attempt of this kind has never been undertaken. Such instruments as exist do not, in the committee's judgment, deal with these two goals with sufficient directness. At every step in the development of these instruments, therefore, the committee has had to break new ground.

4.3 Populations to be Tested

We will study IEA Populations II and IV: all 14-year-olds, and all students in the final secondary school year. In addition, students in Population IVS, who are specializing in literature, will be asked to write essays in response to a short story by Hemingway. IEA Population I will not be included in this study. In most countries there is no deliberate instruction in literature in the elementary school, from which this population of 10-year-olds would be drawn.

It should be emphasized that the collection of essays from Population IVS is considered by the committee to provide a very valuable collection of data, which can be used not only for the present study, but also for future studies of the nature of literary responses among young people. This collection of student essays will form a kind of data bank for such studies during the years ahead.

4.4 The Instruments to be used

As finally developed, the instruments deal primarily with the nature and the quality of the student's response to some actual texts, with his preferred response (as related to "Elements of Writing about Literature") and with his attitude towards literature.

Practical considerations led the committee to develop a set of measurement tasks that can be completed by most students in two hours. Accordingly, the following agreements were ultimately developed:

1. Each student will read and respond to two short stories.

2. A total of four short stories will be used. One of these will be a "common test," administered to all students in both populations. The other three texts will be rotated among studen in Populations II and IV.

3. Two other kinds of instruments were required: an inventory of attitudes towards literature, and a "response preference" test, based on Professor Purves' "Elements."

4.5 Preparation of Tests

The development of the actual tests required approximately one and a half years. Since the nature of the IEA project made it impracticable to read and analyze essays written in response to texts, the committee considered alternatives. Professor Britton, of the Committee, had developed a technique of linked open-ended questions, which led the student through some elements of a given text, then asked that he interpret them. These open-ended questions seemed highly valid, while also corresponding closely with typical kinds of instruction in classes. However, such responses gathered from large populations internationally would not be manageable.

The committee therefore undertook a "crossover" study, in which the question at issue was whether multiple-choice items could elicit essentially the same responses from students as the open-end questions in which the students composed their own responses.

The crossover study (see Appendix F-3) was carried out between January and June, 1968, based on samples of approximately 400 Population II and IV students in England, and the same numbers in the US. The principal findings of this study were that the two test forms - open-ended and multiple choice - correlated at approximately +.70 which was high enough to justify proceeding with the multiple choice form.

Some other questions also arose. Given the assumption that responses to literature were at the heart of the matter, what literature should the IEA students respond to? Clearly, if a way could be found to manage it, lyric poetry should be included, for response to lyric poetry is the most "literary" of responses.

The most obvious problem in international testing of responses to lyric poetry is that of translation. A lyric poem in translation is actually a new poem created on the basis of the original. No one knows what changes, both obvious and subtle, have taken place. Since we were to deal with languages and cultural differences as diverse as Farsi, Italian, and American, it was apparent that translated lyric poems could not be treated like a common text.

What of equivalent lyric poems? Could poems from similar periods and with similar themes, be used? To examine this possibility, a small comparative study was undertaken (see Appendix F-3). The conclusion, not unexpected, was that understanding of literature was text bound, so that more than one text would be needed to indicate a student's general understanding, but that two texts, no matter how similar on the surface, were not similar enough to provide comparability. Regretfully, the committee abandoned further attempts to deal with cross-national, cross-language comparison of responses to lyric poetry.

The experience of the study participants to this point had, however, made it evident that prose passages can be translated with enough equivalence to permit fruitful comparisons of responses. The committee therefore selected from a long list of possibilities, six prose narratives, five short stories, and a selection from an autobiography. The selections were:

My Childhood, by M. Gorki (Russian), a selection

The Use of Force, by William Carlos Williams (U.S.A.), a short story

Just Lather, That's All, by Hernando Tellez (Spanish), a short story

The Sea, by Ana Maria Matute (Spanish), a short story

The Man by the Fountain, by George Hebbelinck (Flemish), a short story

I See You Never, by Ray Bradbury (U.S.A.), a short story

The committee developed both open-ended questions of the Britton type and parallel multiple-choice items for all six of these texts. Not all the participating countries tested all the texts, but enough information was collected to make it possible to be selective of both items and texts.

By October, 1968, the prepared test items had been pretested in Belgium, Chile, England, Iran, Sweden, and the US. Local reasons prevented pretesting in Italy.

Pre-test data indicated for each population in each country and for each item, the following:

a. difficulty (the percent choosing preferred response)

b. discrimination (point biserial correlation of preferred response with total score)

c. percent choosing each of the distractors

d. point biserial correlation for each distractor with total score.

The data thus collected formed part of the basis for decisions to keep, revise, or eliminate items. In addition to this information, the committee had access to the responses of students to Professor Britton's open-ended questions. Wherever possible, the revisions of test items were based on student constructed responses.

Figure 1 illustrates the kind of item analysis information the committee worked with. The item deals with a short story called "The Sea," which will be given to all of the students in the study. In column 1 (DIFF) the percent of students choosing the preferred response is given. In this instance, 96%, 94% and 86% of the students chose the preferred response. The item is too easy.

In column 2 (DISC) is given the point biserial correlations between the preferred response and total score for the students. In this instance, the discrimination of the item is low, but acceptable. Columns A, B, C, and D indicate the percentage of the population choosing each of the options. Option B, "he wants to wash his ears," did not function, nor did option D function well. These two options were therefore revised to make the item more difficult and also to make it more representative of what is important to the understanding of the story. Similar information was collected from all of the countries for all of the items, as has been indicated. In revising the item, the Committee consulted student constructed responses to the same stem in open form. The item as it finally appeared was as follows:

"Which of the following best explains why the boy first said, "I want to see how high the sea will come up on me" (line 12)?

( ) A. He wanted to cover his body with the sea.

( ) B. He wanted to see how tall he was.

( ) C. He wanted to trick his mother.

( ) D. He wanted to show his mother he was not afraid of the sea.

Criteria for the Selection of Items

In selecting items to be included, the committee applied two main criteria, in addition to the usual criteria of indexes of appropriate difficult and discrimination:

FIGURE 1

LITERATURE POP. IV SUB-TEST 3B ITEM B3 TOPIC____ OBJ.___

3. Which of the following best explains why the boy first said, "I want to see how high the sea will come on me" (lines 11-12)?

( ) (A) He wanted to be covered by the sea.

( ) (B) He wanted to wash his ears.

( ) (C) He wanted to trick his mother.

( ) (D) He wanted to hide from the man who cured.

COUNTRY	DIFF.	DISC	A	B	C	D	E	O	NR	COMMENT
Belgium	96	.39	96	0	0	4	-	0	0	N = 45
USA*	94	.36	94	1	3	1	-	0	0	N = 91
Chile	86	.33	86	4	5	4	-	1	0	N = 121
MEDIAN										

COMMENTS

DECISION: ACCEPT___REVISE___REJECT___ FINAL TEST NO. POP.__SUB-TEST___

ITEM____

1. The majority of the items should deal with literary understanding of the text, as contrasted with reading comprehension. By literary understanding is meant ~~the~~ ability to relate part of the work to the whole work and to make reasonable inferences concerning the characterization in a story, the author's employment of metaphor, his selective description of the environment and the like.

2. The items should be true to the text; indirect measures of literary understanding are to be avoided in the text-related items. For example, negative items (e.g., "which of the following four interpretations does NOT fit . . . ") too often appear to depend on general intelligence, or acuity in reading the item stem, at least as much as they depend on literary understanding. Therefore, except in those cases where the negative stem represents the author's original intent, these items were dropped or rendered into positive (declarative) form.

The tests finally developed are as follows:

1. Response preference inventory (20 kinds of responses to each text are offered; the student selects five from these, and assigns a rank to one of these five).

2. Literary interest profile (included in student questionnaire).

3. Attitude towards literature as a field of study (included in student questionnaire).

4. Understanding and Analysis: four texts.

    a. El Mar, by Ana Maria Matute. Eighteen multiple-choice items. A text to be given to all students taking part in the study.

    b. The Use of Force, by William Carlos Williams. Twenty-one items.

    c. De man naar de fontein, by Georges Hebbelinck. Twenty-one items.

    d. I See You Never, by Ray Bradbury. Nineteen items.

Note: Texts b, c, and d are to be rotated between students within a population.

5. The student evaluation of each of the texts he reads, given on a 6-point scale, from +3 to -3.

6. For Population IVS only, an essay in response to The End of Something by E. Hemingway. (The relatively small number of essays so collected will be read and analyzed centrally by trained readers, using the completed code from the "Elements of Writing about Literature." They will be kept for use by the international research community.)

The information to be obtained in this study may be classified into input, process and output variable, with the understanding that some may function both as process and as output. The measure of literary understanding, of attitude toward literature, of literary preference, and of response preference may all be considered as output of instruction in literature, since they reflect important goals of instruction in this field. As is true in all the subject areas in the IEA study, such variables as sex, age, rank in class, socio-economic status of the family, membership in one or another ability group or stream will be related to these output variables. Similarly, the relation between the dependent variables and such environmental variables as the kind of instructional materials used, the length and frequency of class periods, the kind and amount of independent work (including homework), the frequency and nature of external examinations, can also be examined.

Perhaps more interesting than these is the possibility of subdividing the population according to patterns of literary understanding (taken as sub-scores), seeking to predict attitudes, school environment and other matters.

It is intended that the data be analyzed to reveal patterns of literary understanding and behavior, as contrasted with simple scalar measures. For example, the responses to the texts can be examined for patterns of error as well as for patterns of success. Broad human themes are treated by these texts: love, death, aging, violence. The treatment of these themes by students responding to the texts will have very considerable intrinsic interest, being related to country, school type, age, sex, attitudes, etc.

One interesting consequence of the work so far has been the publication of some aspects of the research that have gone into the preparation of these tests. "The Elements of Writing about Literature" by Professor Purves and Victoria Rippere was published as Monograph No. 9 in its research series by the National Council of Teachers of English (USA) in 1968. This monograph has been translated into French for publication in 1969 in Paedogogica Experimentalis in Belgium. The crossover study described earlier will be reported by Dr. Bruce Choppin in

Vol. XV No. 2 of the International Review of Education in June, 1969. Several participants in the study report that they have been called upon to discuss the technical problems and possibilities of the research at professional meetings during the past two years. The study appears to be attracting a considerable amount of attention.

Booklet 3 in Volume II includes the final Literature tests.

CHAPTER 5

French as a Foreign Language

5.0 Introduction

The investigations and discussions leading to the decision to include French as a Foreign Language in the I.E.A. study are described elsewhere in this report. (See Chapter I and Appendix A5.)

The original members of the International Committee on French as a Foreign Language, appointed in 1966 were:

Prof. J. B. Carroll (Chairman)	(Formerly) Harvard University Graduate School of Education, Cambridge, Mass. (From Feb. 1967) Educational Testing Service, Princeton, N.J.
Dr. John L. D. Clark	Educational Testing Service, Princeton, N.J.
Mrs. Clare Burstall	National Foundation for Educational Research in England and Wales, London, England
Prof. Guy Capelle	(Formerly) Dept. of Romance Languages, University of Michigan, Ann Arbor, Mich.; Associate Director, Center for Research in Language and Language Behavior, University of Michigan, Ann Arbor, Mich.
Prof. Wallace E. Lambert	Dept. of Psychology, McGill University, Montreal, Canada
Dr. H. N. Simai	National Organization of Teacher Training and Educational Research, Tehran, Iran

At the initial meeting of the International Committee, held in Hamburg, in November, 1966, Professor Lambert was represented by Dr. John Macnamara, University of Dublin. In 1967, Dr. Erland Kruckenberg, Lärarhögskolan, Stockholm, Sweden, attended a meeting of the Committee.

5.1 Preparation of Rationale

The first task of the International Committee was to develop the general design of the contemplated studies, detailing hypotheses, variables, and target populations. Following are some of the main points agreed upon at the initial meeting:

1. It would be desirable to have testing done at various educational levels, ranging from a population at the 10-year-old level, to

the pre-university level. In including the 10-year-old population, the Committee was cognizant of the considerable interest that had been evinced in foreign language teaching at the elementary school level. For example, at two international conferences on foreign language teaching at the primary level, that had been sponsored by the UNESCO Institute for Education (see Stern, 1963; Stern, 1968), it had been recommended that international studies of foreign language achievement at the primary level should be conducted.

2. An attempt would be made to develop uniform tests of French proficiency at the various levels, including tests of listening, reading, writing, and speaking.

3. The major emphasis of the studies would be on the evaluation of actual proficiency in the French language, in both spoken and written aspects. However, to the extent possible, some consideration would be given to the evaluation of attainment in "cultural" and "general educational" objectives. (Subsequently, it appears to be impracticable to include separate measures of these objectives in the testing program in view of time limitations).

4. The Committee was of the opinion that in view of the fact that in many of the countries to be surveyed French is only an optional subject, the study could better be directed to the investigation of variables involved in yielding high proficiency rather than to the investigation of sheer amounts of "national yield." The comparison of several countries with respect to conditions under which good or poor results are achieved would, it was felt, make for more interesting and useful conclusions than any conclusions based on simple cross-national comparisons.

The conclusions reached by the International Committee at its November 1966 meeting were communicated to National Centers (see Appendix G-1). Voluminous replies and comments were received from many of the National Centers. Many National Centers remarked that it was impossible for them to state any fixed list of "national objectives," there being a variety of objectives represented in different courses and levels of instruction in their countries. In general, however, the National Centers approved the outlines of the proposals made by the International Committee.

5.2 Preparation of Materials

By July 1967 enough information had been received from National Centers to make possible a more precise formulation of the testing program. Accordingly, a sub-committee of the International Committee met in Princeton, N.J., July 31st - August 4th, 1967 (Carroll, Burstall, Clark). The kinds of tests proposed by this sub-committee for the various populations were as follows:

Population One (Age range: 10.0 - 10.11)

Listening Comprehension: Primarily tests in the "four-picture" format wherein the child has to indicate which of four pictures is referred to by a spoken utterance he hears on tape. This format eliminates the problem of variations in reading comprehension ability and therefore tends to be a test of pure listening ability.

Reading Comprehension: Mainly tests using the four-picture format, but with the stimulus being a short printed sentence.

Speaking: The test would consist of three sections: Pronunciation (child imitates spoken utterance and is judged for accuracy of pronunciation); Structural Control; child answers simple spoken questions based on pictures; Fluency: child asked to describe in some detail two or three pictures containing a variety of situations and elements.

Populations Two and Three

Listening Comprehension: Mainly, multiple-choice questions involving spoken options. (Later, it was decided that the options would be printed in French.)

Reading Comprehension: Printed passages (of various types) followed by printed multiple-choice options - all in French.

Writing: Various types of constructed-response items.

Speaking: Similar to the test for Population 1.

Population Four: (Pre-university grade. Both a "survey" test and an "advanced" test would be developed - the survey test for all students to be tested in the IEA sample, and an advanced test for students specializing in French. (Population IVS.) Formal tests of listening, reading comprehension, and writing would be constructed. A speaking test would also be made available.

A start was made on developing all these test forms required. The work on tests for Population 1 was based primarily on a series of tests that had already been prepared by Mrs. Burstall for a project she was working on in England. The test materials for the other populations were for the most part drawn from non-secure item files at Educational Testing Service. Some of the materials supplied from National Centers were also found useful, and a number of new test items were developed by Dr. Clark and Mrs. Burstall.

The second full meeting of the International Committee took place in London, October 9-13, 1967. The members present reaffirmed the Rationale formulated at the November 1966 meeting but took note of certain revised lists of hypotheses developed by several countries.

The committee at this meeting also thoroughly reviewed the draft tests that had been prepared by Mrs. Burstall and her staff (in England) and Dr. Clark (in the USA). Various detailed points were commented on, and a number of major deficiencies in the tests were pointed out. In several days of work subsequent to the committee meeting, Mrs. Burstall and Dr. Clark were able to correct most of the deficiencies (e.g., overemphasis on "rural" content in some of the Population I tests).

Further, the committee addressed itself to the construction of attitudinal and descriptive scales that would be administered along with the tests, and drafted instructions to the National Centers regarding the administration of the Tests, translation of instructions, etc., for the pre-test phase of the project.

The tests in the form in which they were prepared just subsequently to the October 1967 committee meeting were sent out to National Centers for further comment. A number of detailed suggestions were received, mostly having to do with phraseology of the test items, and were taken into account in the preparation of the final forms for pretesting that were then sent out to National Centers in March 1968 along with complete instructions for administration.

The test rationale and description of pre-tests is included as Appendix G-2.

5.3 Pre-Testing

A description of the tests sent out for pre-testing is as follows:

Population I:

Listening: Two forms were constructed, L1 and L2. Each consisted of 5 sample items and 50 test items (there was some overlap of items between the two forms). The items were in a "four-picture" format whereby the child had to listen to a spoken utterance on a tape and indicate which picture it referred to. Time for administration: about 30 minutes.

Reading: Two forms, R1 and R2, with some overlap between the forms. Each test had 5 sample items, followed by 35 "four-picture" items where the child had to indicate which picture a printed stimulus referred to, and 10 multiple-choice items based on short printed texts. Suggested time for administration: 30 minutes; however, the tests were to be administered without rigid time-limits.

Speaking: One form, S1, consisting of three sub-sections:

Pronunciation: 21 sentences recorded on tape; the child was to imitate each sentence, his response being recorded on tape for later evaluation of his pronunciation of specific phonemes or intonations.

Structural Control: 10 sample pictures; the child was to make a one-sentence response answering a question (in French) about what was shown in the picture.

Fluency: 3 somewhat more complex pictures; the child was to choose one of the pictures to describe in French.
The Speaking test was to be given to only a sample of children, roughly divided among "good," "average" and "poor" learners. Estimated administration time: 15 minutes, individual administration.

Population II:

Listening: 10 four-picture items; 35 four-choice items with options printed in French, based on spoken (recorded) texts of varying length. Administration time: about 25 minutes.

Reading: 45 multiple-choice items, options printed in French, based on short sentences or paragraphs. Suggested administration time: 35 minutes.

Writing: 28 items of the "fill-in" type, where a word had to be supplied from the context.
5 items requiring the "correct French reply" to questions, usually involving change of person, use of the negative, use of pronouns, etc.
8 items involving rewriting sentences to change number, person, etc.
1 item involving writing of a 6-line "directed" dialogue.
Suggested administration time: 35 minutes.

Speaking: Test S2. Individual administration, about 15 minutes per pupil.
Four parts:

Pronunciation: (Same as corresponding section of Test S1.)

Structural Control: Similar to the corresponding section of Test S1, but with different questions and pictures.

Oral Reading: Pupil reads aloud a French paragraph.

Fluency 1: Pupil chooses one of three picture-sequencies, to recount the story implied by the pictures.

Fluency 2: Pupil chooses one of three pictures showing certain situations; he is to tell, in French, "what has happened," "what is happening now," and "what will happen next."

Population IV: (Survey tests)

Listening: 12 four-picture items
50 four-choice items with options printed in French.
Administration time: about 30 minutes.

Reading: 53 four-choice items, options printed in French, based on short passages in French. Suggested administration time: 40 minutes.

Writing: (There was no special writing test for this population; where one was administered, it was usually that for Population IVS.)

Speaking: (There was no special Speaking Test for this population; where one was administered, it was that for Population II.)

Population IV (Specialist)

Listening: Test IVS. 5 four-picture items, plus 46 four-choice items with options printed in French. Administration time: 30 minutes.

Reading: 44 four-choice items based on passages, dialogues, etc. of varying lengths. Suggested administration time: 35 minutes.

Writing: 53 "fill-in" items; 7 "correct reply" items; 7 items requiring grammatical changes in sentences; 2 "directed essays": student was to write an essay using certain suggested words and phrases. Suggested administration time: 35 minutes.

The revised definitions of target populations are as follows:

Population I: All students aged 10.0 - 10.11 at the end of the school year who will at that time have completed two or more school years of instruction in French (whether or not they are still studying it).

Population II: All students aged 14.0 - 14.11 at the end of the school year who will at that time have completed two or more school years of instruction in French (whether or not they are still studying it).

Population III: All students at that point of secondary schooling where there is a major drop-out from French (or from school altogether) who will at the end of the school year have completed two or more school years of instruction in French (whether or not they are still studying it.

Population IV: All students in the pre-university year (approximate grade level: Grade XII) who at the end of the school year will have been studying French for two or more years (whether they are still studying it or not).

Population IVS: All students in the pre-university year who have studied French for at least three years and are still studying it. (National Centers are to make operational definitions of this target population for their own countries. These students are those, in the country, who are regarded as specializing in French. The operational definitions are to be sent to the IEA coordinator.)

An indication of the quantity of pretesting is given here:

2

Population I tests:

Listening L1	Reading R1	Speaking*
England (N=135)	England (N=152)	England
U.S.A. (N=102)	U.S.A. (N=102)	U.S.A.
Scotland (N=28)	Scotland (N=29)	
Listening L2	Reading R2	
England (N=150)	England (N=138)	
U.S.A. (N=138)	U.S.A. (N=137)	

(Note: Also, Listening L1, L2 and Reading R1, R2 were given to a sample of 160 French children, 80 children took tests L1 and R1; 80 took tests L2 and R2. This was to test for possible ambiguities in the items.)

Population II tests:

Listening	Reading	Speaking*	Writing*
England (N=120)	England (N=117)	England	England
Iran (N=91)	Iran (N=102)	Scotland	Iran
Scotland (N=149)	Scotland(N=149)	U.S.A.	Scotland
U.S.A. (N=155)	U.S.A. (N=154)		U.S.A.

Population IV tests:

Listening	Reading	Speaking*	Writing*
Chile (N=159)	Chile (N=169)	Chile	England
England (N=119)	England (N=126)	England	Scotland
Iran (N=100)	Iran (N=99)	Scotland	U.S.A.
Scotland (N=144)	Scotland (N=150)	U.S.A.	
Sweden (N=141)	Sweden (N=131)		
U.S.A. (N=95)	U.S.A. (N=147)		

Population IVS tests:

Listening	Reading	Speaking	Writing*
Chile (N=59)	Chile (N=80)		Chile
England (N=77)	England (N=70)		England
Scotland (N=88)	Scotland (N=93)		Scotland
U.S.A. (N=220)	U.S.A. (N=218)		U.S.A.

The objective tests were item-analyzed at the National Centers according to instructions that had been sent out by IEA. The Speaking and Writing tests were scored at the National Centers, but item-analysis procedures were not applied. Instead, samples of test responses (usually, a one-tenth sample of the tested population) were forwarded to Hamburg for evaluation by the International Committee.

---

*Tests results for Speaking and Writing were received for only small samples of the tested populations.

A meeting of the International Committee was held in November 25 - 29, 1968 to consider the item analysis results and prepare semi-final forms of the tests. Although many of the test items were found to be effective in terms of difficulty and discrimination indices, many others had to be revised or rejected altogether. Nevertheless, it was possible to develop semi-final forms of the Listening and Reading tests for all populations. On the basis of the limited results available for Speaking and Writing tests, it was decided that fairly extensive revisions for these tests were needed, necessitating the prospect of another round of pre-testing to be done in the spring of 1969. In fact, it was thought desirable to have another round of pre-testing for all the tests in view of the revisions that had been made.

5.4 Preparation of Penultimate Versions of Tests

The semi-final forms of tests constructed at this meeting may be described as follows:

Population I

Listening: 40 four-picture format items; 20 minutes working time plus administration time, stimuli on tape.

Reading: 40 four-picture format items, stimuli printed in French; 25 minutes working time, plus administration time.

Speaking: Three parts, as before:
Revised pronunciation test
Revised structural control test
Fluency: revised marking scale and instructions for its use. Revised pictorial material.

Population II

Listening: 40 items (some of the four-picture type; others with options printed in French); 20 minutes plus administration time.

Reading: 40 items; 25 minutes plus administration time.

Speaking: Pronunciation (revised)
Structural control
Oral reading (shortened paragraph)
Fluency: Revised marking scale.
Total administration time: about 15 minutes.

Writing: Section 1; 27 one word completion items.
Section 2; 8 multiple completion items (20 points)
Section 3; free composition passage.
Total administration time: about 35 minutes.

Population IV

Listening:	45 items; 25 minutes, plus administration time.
Reading:	40 items; 25 minutes, plus administration time.
Speaking:	(same test as for Population II)
Writing:	Section 1: 16 one word completion items. Section 2: 12 multiple completion items (37 points). Section 3: free composition, two passages. (a few additional items will also be included in Sections 1 and 2). Total administration time: about 40 minutes.

Population IVS

Listening:	45 items; 25 minutes, plus administration time.
Reading:	45 items; 30 minutes, plus administration time.
Speaking:	(Same as for Population II.)
Writing:	(Same as for Population IV.)

Listening to the Speaking test tapes that had been sent in and comparing the responses to the marks that had been assigned locally, the members of the committee judged that the marking was too variable across countries to continue a policy of local scoring. Therefore, it was decided to set up, in the future, a central scoring facility where raters could be trained to more or less uniform marking standards across countries. Provision will be made so that each Speaking test sample will be marked by three independent raters. The same will apply to Section 3 of all Writing tests.

The tests presented in Volume II, Booklet 5 are the pre-test versions.

5.5 Timetable for Future Activities of the Committee

Dec. 1968 - Jan. 1969:	Preparation of second pre-test forms, tapes instructions, etc., suitable for sending out to the participating countries.
Feb. - May 1969:	National Centers prepare for pre-testing; item analyses to be sent to Hamburg by the beginning of June.
May - June 1969:	IEA Secretariat and Central Scoring Committee prepares materials for a summer meeting of the International Committee.
June - July 1969:	Meeting of the Committee to review second-round of item-analysis results, review results on student and teacher questionnaires, attitude and exposure scales, and refine lists of hypotheses and variables. (Tentative dates and venue of meeting: Nadrin, Belgium, June 27th - 4th July)
July - Dec. 1969:	IEA staff prepares materials for "dummy-run" testing in Spring 1970 (answer cards, instructions, etc.)

CHAPTER 6

ENGLISH AS A FOREIGN LANGUAGE

6.0 Introduction

The original members of the International Committee for English which first met by invitation of I.E.A. at Hamburg in November 1966 were

E. Glyn Lewis (Chairman) H.M.I. and later Director of Educational Research at University of Wales, Swansea

Mrs. Heitner, Israel Department of Education

Monsieur Morette, Senior Master at the Lycee Henri IV, Paris

Dr. Agnes Niyekawa-Howard, University of Hawaii

Mr. R. Piirtola, University of Jyväskylä, Finland

Professor C. H. Prator, U.C.L.A., United States of America

Subsequently, Mr. Ian Dunlop of the British Centre, Stockholm was invited to join the Committee. He has participated in the work of the Committee by correspondence from the beginning.

6.1 Preparation of Rationale

The Chairman presented the Committee with a paper outlining provisionally the variables which might have to be taken into account in designing the study in line with the intentions of I.E.A., and this was accepted with some modification at the first meeting. In the light of the briefing the Committee had received, it set out a series of 24 hypotheses which could form the guidelines for the study, and indicate at the same time the kinds of research instruments tests and descriptive measures which would have to be designed. It was agreed that only two major target populations should be investigated - I.E.A. Populations II (and III), and IV. It was decided that Population I, namely the 10-year-olds would not provide a sufficiently rewarding and satisfactory sample in the countries likely to participate in the study, since too few children would have been taught English long enough to justify testing at that age.

The Committee prepared a provisional 'content analysis' formula to be submitted for completion by the National Centers, and a second formula for completion relating to the objectives of English teaching in each of the participating countries. The areas of skill and knowledge to be studied were also discussed, and it was agreed that no attempt would be made to test knowledge of the history and cultural background of English speaking countries, and that the achievement tests would be directed solely to investigating the control over the language possessed by the students. It was agreed also that the tests of proficiency should take into account as wide a range as possible of actual achievement in view of the wide currency of the English language and the fact that in some countries it was the first foreign language and taught to a high proportion of the school population. Finally at the first meeting some consideration was given to the form the tests should take. It was agreed that of the two alternative approaches to the preparation of these tests - namely, first to design the tests according to the linguistic level relevant at that point such as vocabulary, grammar or the sound system of the language; or second, the language skills possessed by the student namely listening, reading, speaking and writing, the Committee would adopt a consistent approach and follow the second criterion and so prepare listening, reading, writing and speaking tests ensuring at the same time that the important linguistic aspects were satisfactorily covered according to the content analysis agreed upon by the Committee. It should be emphasized that several of the initial decisions have had to be reviewed (and in some cases revised) by the International Committee in the light of the views expressed by National Centers and the experience of attempting to implement those decisions. For this reason a section of this report is devoted to a consideration of some of the more interesting issues that have had to be resolved.

Following the first meeting, the National Centers were informed about the preliminary thinking of the International Committee. They were invited to comment upon the approach suggested and to supply information concerning the content analysis of the curriculum and the teaching objectives which they entertained. By and large the response of the National Centers proved encouraging, although one or two questioned the possibility of creating uniform international tests, and the advisability of using multiple choice techniques.

6.2 Preparation of Materials - First Stage

The second meeting of the International Committee considered the replies from National Centers and decided on the proportion of testing time to be devoted to the various skills. It was agreed that the greater proportion of time should be given to reading and writing items, and that speaking tests should be restricted to a sub-sample of the sample of the relevant target populations. The International Committee also considered a paper which collated the content analyses submitted by the National Centers, and a second paper which collated their statements of teaching objectives. The Committee modified both papers slightly and then accepted a collated content analysis on which the preparation of the actual test items would be based. They accepted also a formula of teaching objectives which was to provide a basis for the proportionate emphasis to be placed in each target population upon the aspects of the English language to be tested. This information was conveyed to the National Centers who were also informed that

Listening Tests	would be multiple choice tests involving oral written responses. The whole test would be in English.
Reading Tests	would include items of varying length covering all aspects of the language - vocabulary, structure, etc., with multiple choice options in English exclusively.
Writing	would be tested incidentally by means of the reading tests and the written responses to the listening tests.

The same considerations would apply to Population II as to Population IV, save that in the latter case passages of greater length and complexity would be introduced into the Reading tests. It was decided to delay consideration of the form of the speaking tests until greater information had been obtained from I.E.A. concerning the nature of the sub-sample to be tested. The preparation of these tests was left to the Chairman in the first instance. It was also decided that there should be a considerable overlap between the two major target populations so far as test items was concerned, and that this would meet the needs of the "Survey" test component of Population IV. At the same time it was agreed that a considerable part of the Population IV tests should be of a quite advanced standard. With all this in mind the National Centers were asked to supply as large a selection of test items thought suitable in their countries from which a selection f the first draft test batteries could be prepared.

The National Centers responded with useful comments and very comprehensive selections of test items for the relevant target populations. The Chairman then prepared a provisional battery of tests for each population based upon the selection offered by National Centers, published and non-restricted materials in Britain, and material supplied from non-restricted items in the possession of the English Language Center at Ann Arbor, Michigan and the English Language Center at Georgetown University. These batteries were then submitted to the National Centers and members of the International Committee for comment. They were returned with useful corrections and comment and the Chairman revised them in the light of these comments. Finally they were discussed and refined by the Chairman in consultation with Mr. Piirtola at a series of sessions at the National Foundation for Educational Research, London. Following the final revision, the tests were submitted to two groups of thirty children of appropriate ages in schools in Wales where the English language is not the mother tongue of the majority of the children. This was done to test for gross ambiguities and unsuspected errors or difficulties. Very few of these were discovered. Consequently the test items were made up into 8 units of testing material for Population II and 10 units for Population IV. Each unit consisted of no less than three test items and not more than 45. Test items were repeated in some instances in several units so that a comparison between the response of various pre-testing contres might be established. The pre-test units were then submitted to various National Centers for pre-testing.

Breakdown of Pre-tests

The following is a breakdown of the composition of the tests prepared for pre-testing -

Population II

Listening

i Dictation.
ii Phonemic identification tested by picture and multiple choice option of taped words.
iii Listening comprehension tested by taped short statement or series of statements and multiple choice response.

iv Recognition of sounds - differences and identity - by means of a series of three taped words.
v Recognition of differences or identity of intonation in three taped sentences identical in structure but sometimes different in intonational pattern.

Reading

i Four different types of tests relating to pronunciation, spelling and stress.
ii Eleven tests of structure control including the use of some tests involving pictures. These tests cover all the aspects included in the content analysis agreed upon, including especially word order and the ability to transpose direct and indirect speech, active and passive, question and statement.
iii Vocabulary - eleven different types of tests some of these being incidental to listening and reading, but some specifically designed for vocabulary alone - namely tests of definition, synonyms and antonyms in and without context.

iv Recognition of meaning was tested partly by the identification of narrative or descriptive detail in continuous passages or by means of the recognition of correct inference or incidentally in the tests listed above.

Population IV

Listening

i Dictation.
ii Recognition of intonation patterns combined with differences of emphasis.
iii Listening comprehension as Population II - iii.
iv Listening comprehension tested by taped short conversations and with choice of three written responses.
v As for Population II - iv.
vi Listening comprehension involving recognition of key phonemes for discrimination of meaning.
vii As for Population II - v.
viii Listening comprehension based on choice of correct response to fairly simple taped statements.
ix Identification of correct stimulus sentence given the correct written response. The identification from among three stimulus sentences depends upon the correct relation of intonation and emphasis to meaning.
x Straightforward listening comprehension of details of description and narrative.

Reading

i Tests on recognition of correct stress related also to the differentiation of nouns and verbs.
ii Grammar and structure - These are very largely as for Population II in form though not in content. More emphasis on transform tests, and on the use of extended passages for the recognition of details of correct usage. These tests are basically miscellaneous in character; they do not deal with categories of structure and usage in blocks of items. Nevertheless each test item is easily identifiable for analysis in terms of the element it tests.
iii Vocabulary - Very largely the same in form as for Population II.
iv Style - Tests have been included which seek to investigate performance in recognition of appropriate registers and styles. These tests take the form of linking disparate statements to form a coherent passage or paragraph. For each possible statement, a multiple choice of styles or registers is offered.
v Comprehension - Several tests of comprehension are provided - some devoted to the ability to draw correct inferences, to come to correct conclusions, to recognise possible implications, as well as straightforward recognition of descriptive or narrative detail.

Speaking and Writing Tests

i Continuous oral composition subsequent to provision of eight stimulus items.
ii Oral description and narrative, based on series of pictures.
iii Mimicry-repetition of spoken sentences.
iv Oral response to specific taped questions referring to taped short statements.
v Oral response to specific taped questions referring to taped short statements.
vi Single word taped response to simple pictures of single objects or actions.
vii Written completion of sentences. This test is to some extent open-ended since what is required is *a* correct response rather than *the* only correct response possible.
viii Continuous written composition, including letters, based upon "guided composition principle."

6.3 Pretesting

The following pre-tests were administered:

Population II:

Listening		Reading	
Germany	(Test 11)	Germany	(Tests 1,2,3,4,5,8)
Finland	(Test 12)	Finland	(Tests 5,7,8,9,10)
Iran	(Test 11)	Iran	(Tests 1,3,6,7,9,10)
Italy	(Test 12)	Italy	(Tests 2,4,6,8,10)
Netherlands	(Test 11)	Netherlands	(Tests 1,3,4,5,7,9)
Sweden	(Tests 11 & 12)	Sweden	(Tests 2,4,6,8,10)
Thailand	(Test 12)	Thailand	(Tests 1,2,3,5,7,9)

Population IV:

Listening		Reading	
Chile	(Tests 9,10)	Chile	(Tests 1,2,3,4)
Germany	(Tests 10,11)	Germany	(Tests 4,5,6,7)
Finland	(Tests 11,12)	Finland	(Tests 1,2,7,8)
Iran	(Tests 9,12)	Iran	(Tests 3,5,6,8)
Italy	(Tests 10,12)	Italy	(Tests 1,2,3,4)
Netherlands	(Tests 9,11)	Netherlands	(Tests 4,5,6,7)
Sweden	(Tests 9,10)	Sweden	(Tests 3,5,6,8)
Thailand	(Tests 11,12)	Thailand	(Tests 1,2,7,8)

6.4 Attitude and Other Descriptive Measures

Meanwhile the International Committee with guidance and assistance from the officers of I.E.A. gave attention to the preparation of attitude scales and descriptive measures. Among the descriptive measures the English Committee had decided upon are the following:

i A method analysis.
ii A measure of linguistic distance.
iii A description and measure of linguistic background of students.
iv A description of the linguistic situation in each participating country.

The meetings of the International Committee as a whole did not provide the time necessary to develop these instruments. Consequently the Chairman took advantage of a visit to the United States for other reasons to consult with Dr. Niyekawa-Howard on the preparation of these measures. After two days of intensive consultation tentative proposals were made to the officers of I.E.A. These were further refined in discussion between the Chairman and officers in London and in collaboration with Mrs. Burstall of the French International Committee at the Stockholm meeting of I.E.A. in December, 1967.

6.5 Points of Principle Raised During the Study and Justification of the Committee's Approach

(a) An instrument for the analysis of method

The English Committee has no desire to exaggerate the importance of difference of method in helping to determine levels of achievement. Nevertheless the Committee has always wished to make some attempt at such an analysis, and to relate it to other variables. The only question to undertake this specifically for language is as follows:

> Indicate by checking in the appropriate column the amount of importance in general of each of the following in learning English/French as a foreign language.

	Not Important	Important	Very Important
(a) Listening			
(b) Speaking Correctly			
(c) Speaking Fluently			
(d) Understanding Spoken English/French			
(e) Use of Formal Grammatical Drills			
(f) Grammatical Explanations			
(g) Reading for Clear Understanding			
(h) Reading Extensively			
(i) Writing Correctly			
(j) Writing Fluently			

We wish to include other questions which would relate to the teaching of language specifically, since we are convinced that questions relating to teacher-pupil relationship and to teaching method in general terms are more than likely to miss the relevant and significant aspects we wish to analyse.

(b) Measure of language distance

The problem of negative or positive transfer from the act of learning one language to the act of learning another is important, simply because the transfer will be facilitated not so much by the qualities possessed by the student or by the teacher, or the system under which the language is taught as by the relationship which may exist between the two languages - the distance of the two languages. This operates at all levels, vocabulary, the sound system and the grammar. One way of taking this factor into account would be to use a contrastive analysis of the mother tongue and English. But apart from the fact that this process would be enormously expensive in terms of time and money and would hold up the preparation of the achievement tests for a considerable time, it would not solve the problem of preparing a set of tests which have to be administered across the board - a set of tests to be applied to all countries in common. A contrastive analysis would enable us to prepare tests which were fair to each country separately since they would take each native language into account. But it would not enable us to provide the basis for comparison of achievement.

The procedure adopted by the English Committee is not entirely or even broadly fair to all countries, but if we can include a measure of distance, it can be said to be equally unfair and the differences in unfairness can be partialled out. The procedure is to ask the teachers first of all whether they speak another foreign language and if they do to indicate whether the learning of the two languages, in their view is equally difficult or more/less difficult. We ought to get a comparison of the difficulty of learning English expressed in terms of the difficulty of learning several other foreign languages, and this information will enable obtaining a measure of the distance of English from each of these languages. It is accepted that this measure is subjective, but it will be based on the best judgment of experts. But in order to refine this rough instrument a little we would like to include the questions which refer to this issue in the questions which may be addressed to National Centers with a request that the National English Committee in each case should answer it.

A second approach to the assessment of linguistic distance is in the Teacher Questionnaire. This is an attempt to identify fairly precisely the half dozen or so most important areas of linguistic interference, or differences between the mother tongue and English. We wish to correlate in the best way possible the results from this question and the error analysis of the achievement tests in their final form. Two results would be obtained in this way: a check on the a priori theoretical contrastive analysis given by the teachers in reply to the relevant question. And second, in so far as the correlation was positive and significant, it would be possible to explain the incidence of errors which occur.

(c) Which English?

The various National Committees were asked to indicate whether they preferred the British or American variety of English where the differences between the two might be important. This was done because some of the participating countries are likely to be under the influence of one rather than the other. In the event the National Committees who were addressed decided in favour of British English. Since that time one or two other countries have expressed a wish to participate, e.g., Thailand, and it may be that such countries might prefer American English.

(d) Selection of vocabulary and structures

English is not favourably placed when it comes to decide upon the criteria of selection to be adopted in an international test. A good deal of very good work has been done in this area in the U.S. and Britain but simply because, unlike French, there has never been an official support for one set of criteria or for a course employing a stated set of criteria (as in Francais Fondamental), the teaching of English as a foreign language in different countries reflects the acceptance of several different criteria. If we adopted one set of criteria countries using it would be advantaged compared with countries who based their courses on a different set. Consequently we decided to work pragmatically and empirically, using the broad experience of the International Committee, corrected by the feed-back from National Centers. This procedure again is not fair to any country, but it has the advantage of being equally unfair to all.

(e) Grading of vocabulary and structures

The steepness of the grading of the tests has been a considerable problem since, to be satisfactory, every topic tested had to be represented several times and this makes for a lengthy test. If in addition we decided on a too gradual grading, the tests would become impossibly long. It was decided to have a very lengthy set of pre-tests allowing for repetition of types, and gradual grading. The final tests will be within the time limits specified by I.E.A. and within these limits the grading will be as gradual as possible.

(f) Multiple choice?

The English Committee decided on this form of testing without any considerable discussion, partly because it has proved a successful method of setting out language tests in Britain and the U.S. aimed at foreign students of English; and partly because the nature of the present exercise and the resources available made this form of test very convenient. Nevertheless one National Center has expressed its reservations about the appropriateness and even validity of this kind of testing when applied to English.

(g) Reflection of differences in national objectives for English

When National Centers were asked to state the objectives pursued by their own teachers of English, there were some considerable divergences, and it has been a problem to try to include sufficient tests at each of the target population levels to meet the varying aims. This matter mainly affects the balance of the tests - the relative weight to be given to the varying skills which are tested, some countries favouring reading as the main objective, while others favoured speech. Others again favoured translation as a major objective while others discounted it entirely.

(h) Use of the mother tongue in testing

Apart from the instructions, the mother tongue, it was decided, should not be used in these tests. This decision effectively ruled out certain types of tests, especially translation of passages or the interpretation of vocabulary in terms of the vocabulary of the native language.

(i) Cultural bias

Just as there are linguistic problems of positive and negative transfer, so there are cultural problems, or issues of "culture distance." Generally speaking Western European countries are nearer so far as the pattern of their culture goes, to that which the English language is normally related to, than is the case with Middle Eastern or Far Eastern or Asian countries. Consequently the English Committee had to decide on one of two alternatives. First, to relate the content of the tests to the pattern of an English speaking setting. An assumption underlying the choice of this alternative is that all countries participating in this exercise taught English with the objective of participating in an English speaking culture or at least of entering into an empathic understanding of it. This is not a valid assumption since some countries testing English are concerned almost exclusively with English as a vehicular language rather than as a means of culture transmission. The second alternative is to relate the content of the features of ordinary communication. Since this is basic to all teaching of English, the Committee decided in favour of the second alternative. It does mean on the other hand that some countries, e.g., Sweden which places great store on their relating the language to the English "way of life" (and this is becoming increasingly a feature of progressive teaching) are not being given the true assessment of their approach to and teaching of English.

Apart from the problem of culture content as it affects different countries, there is the similar problem as it affects schools within the same educational systems but with different educational orientation. This is particularly the case when we consider vocational as opposed to academic schools. Subjects which might be within the range of coding interest of vocational schools would be outside the range of the academic school and the converse is equally true. This has been pointed out by one national committee (Italy) when submitting observations on the choice of reading passages. In the last resort, unless the list of reading passages is to be extended unduly, the choice of cultural content has to be left to an arbitrary decision, though we hope it is an informed one.

(j) Cognitive aspect of the language tests

It has been accepted all along that any tests of linguistic competence, beyond the most elementary, and excluding the purely imitative, are at the same time an assessment of cognitive processes. Consequently, especially in the tests for Population IV, but to some extent also for other populations, there are several subtests which are tests of the ability to draw conclusions, to infer, and to draw out the implications of the statements made in English.

6.6 Future Action

There have been delays for various reasons, in the pre-testing of the English items. It is expected that all item-analysis will be complete for the next meeting of the English committee in June/July 1969. The tests presented in Volume II, Booklet 5 are the pre-test versions.

CHAPTER 7

CIVIC EDUCATION

7.0 Introduction:

The assessment of children's achievement in Civic Education presents a number of basic problems which highlight the differences between Civic Education and other school subjects.

One of these is the fact that Civic Education is not yet well-established as an academic discipline, so that there is a good deal of disparity between countries in the subject-matter covered and the ways in which it is approached. Moreover, these disparities are more influenced by a particular country's political and socio-economic system than is the case, say, with Science or Mathematics. Thus, in a Monarchy children are likely to learn a good deal more about royal succession and duties to the Crown than in a Republic.

Another difference lies in the fact -- much stressed both in the curricular objectives and by the teachers themselves -- that Civic Education does not merely consist in the transmission of a body of knowledge, but that it aims at inculcating certain common attitudes and values, such as a liberal and democratic outlook, political responsibility, the ideals of tolerance and social justice, respect for authority, and so on. Behind the need to teach pupils certain fundamentals about political, economic and social activities and organisations lies the desire to turn them into effective and responsible citizens of their society. Indeed the cognitive content of the curriculum is frequently used in order to highlight the underlying principles and ideology; thus, information about electoral systems could be utilised to bring out fundamental ideas about equality and majority rule.

One implication of this has been that, as we shall see, the measurement of the children's knowledge (cognitive domain) and the measurement of their attitudes and values (affective domain) has proceeded somewhat separately -- partly because children's attitudes in this sphere so often develop well before they have the relevant knowledge, and partly because affective measures need a rather different type of instrument (attitude scales, projective techniques, and so forth). At a later stage the data obtained through these two sets of measures will be compared and contrasted, and given a theoretical interpretation.

A third difference between Civic Education and other subjects lies in the very considerable influence which out-of-school factors will have on the pupil's knowledge and attitudes. We must expect that the child will be affected not only by what he learns in the classroom, but also by political events and by the entire fabric of the society in which he lives. As far as possible we must, therefore, make an effort to include measures of these among our independent variables.

Finally, we must mention a linguistic and terminological difficulty: in writing about this area, it is difficult to avoid over-simplified labels such as 'patriotism,' 'democracy,' 'equality of opportunity,' 'citizenship' and the like, but we should like it to be clearly understood that such terms are used by us purely for reasons of convenience and not in order to advance any ideological point of view. The labels merely refer to particular curriculum segments or to variables in the conceptual framework; they are defined by their content specifications.

In short, we find that the assessment of achievement in Civic Education cross-nationally is attended by a number of awkward problems: uncertainty about a 'common core,' the need for special attitude measures as well as knowledge tests, strong out-of-school influences, terminology with political overtones creating difficulties in conceptualisation. We have addressed ourselves as best we could to each of these, as will be shown in the succeeding sections.

7.1 Contents of the National Documents:

The members of the International Civic Education Committee of I.E.A. which met in Rome in April, 1967, had before them a series of documents produced by the national centres of W. Germany, Finland, Iran, Sweden, the UK and the USA. These documents contained a digest, for each country, of the curricular aims and contents in Civic Education for different age groups, of textbooks used, and of examination contents. While the country documents differed widely, it was obvious from the start that, in addition to knowledge, all countries laid considerable emphasis on values and attitudes. Here, for instance, is section (1) of article 26 of the constitution of the German state of Bremen:

> "Children should be brought up in a community spirit based on respect for the dignity of all men and on desire for social justice and political responsibility, and should be taught to regard the opinions of others with objectiveness and toleration and work peacefully with other people and nations."

We therefore undertook two detailed content analyses, one dealing with cognitive knowledge, the other with values and other affective variables.

The content analysis of the cognitive domain led in due course to a conceptual framework (see Appendix H-1) which has been the major guiding force in the development of cognitive measuring instruments. The analysis of the affective contents of the country documents showed that, perhaps a little unexpectedly, there was a common core of basic values which are stressed in all six countries concerned. This common core consisted of the political ideology of equality (freedom and civil liberties, tolerance of diversity, majority rule, non-violent conflict resolution, etc.) and citizenship values (such as informed participation, being law-abiding, interest in welfare of fellow citizens, love of country). These values were set out in detail, and were in due course incorporated in the conceptual framework and in the measuring instruments.

7.2 The Development of Cognitive Instruments

Between November, 1966 and September, 1967 the international civic education committee of the IEA study formulated its preliminary test specifications. The first meeting produced a preliminary rationale and model stressing structures of authority. This model was used as an overall conceptual framework which would guide the inquiry until information could be gathered about civic education from documents which national centers in the seven participating nations would produce. National centers were asked to prepare papers which would summarize the major goals of civic education in their country, representative curricular statements, examples of test materials in use, and summaries of textbook materials used in civics or other social studies courses having political content. These reports were received in the spring of 1967 and provided the basis for a subsequent meeting.

Before and during this meeting a content analysis of the national documents also was used to formulate an international master grid (test content specifications) summarizing the major content and ability areas of civic education as taught in the seven countries at the three basic target population levels: the 10 year olds, 14 year olds, and pre-university year students. Each of the cells in these test specifications was given a weight showing the relative emphasis expressed in the natural documents received from the seven countries. Test questions received from the various national centers were then categorized in terms of this master grid. These specifications with the illustrative questions went out for review to the various national centers with a request that new questions be written in accordance with the specifications. When the reviews and questions were received this then provided the committee with the necessary material to produce definite test specifications at a meeting in September, 1967.

7.3 The Civic Education Content Specifications

The results of our review of national center reactions to the preliminary test specifications indicated that terms such as civics, citizenship, or political education were meaningful educational terms (as courses of study or curricula emphasis) in all seven countries. However, these terms had varying degrees of currency in the educational languages of participating nations. In some countries "civics" is treated as a separate subject. In others it is taught through history classes while in still others it is a form of school work emphasized throughout the curriculum. For instance in Sweden, which has formal courses in "civics," topics such as working life and economic problems, international problems, the political process (e.g., public opinion and political parties), law and justice, personal rights and duties, and so forth are taught in "civics" courses. In the United Kingdom, which has little formal training in "civics," these same topics are taught through several subjects.

Our content analysis and survey results revealed that the common core of civics taught in the seven nations dealt with the following major topics:

I. Fundamentals and Nature of Citizenship - citizenship concepts and definitions of terms such as: state, government, patriotism, freedom and responsibilities, rule of law and democracy;

II. Political Processes and Institutions: National Topics such as constitutions, political history, local government, law making, judicial decision making, executive policy making and the civil service, political leadership, public opinion, elections and political parties;

III. Political Processes and Institutions: International Topics such as foreign policy, national defense, comparative politics, the international political system, and international organizations;

IV. Economic Processes and Institutions - topics such as social services, welfare, taxation, labor organizations, and occupations;

V. Social Processes and Institutions - topics such as the family, the school, mass media, traffic, and crime.

Each of these topics was also divided in terms of simple, complex, and abstract operations a la Bloom's taxonomy. In all, approximately 1,500 draft pretest items were submitted to the international committee for possible inclusion in the draft pretests.

7.4 The Civic Education Pretests

On the basis of these content specifications and the draft items, general forms of pretests were prepared for four groups of students, namely the 10-year-olds, 14-year-olds, and preuniversity year populations and a series of tests for those preuniversity year students who have taken extra work in civics, history, and social studies (a specialist group). These pretests (about 850 items) have been or are being administered in each of the participating countries. On the basis of these results the final pretest instruments are being prepared.

In addition to the test analysis information resulting from the pretesting, National Centers have also been providing the civic education committee with feedback information as to whether the content of test questions is appropriate for the various age levels and if the content is suitable in their countries. This has required them to revise test questions in some instances so that the tests will be appropriate.

Two other checks of content validity have been or will be made. First a check list of content in the attitudinal and cognitive domain has gone out to national centers and the results are being compiled. This will aid in the development of the final pretests. What is expected to come out of this testing is a series of subscores for each of the five (I-V) subtopics of the content specifications described above. This will allow us to construct national and international profiles of children's knowledge of international topics, economic and social processes and institutions, citizenship fundamentals, and basic national political topics. Certain other and smaller subscores could be developed showing comparative knowledge about constitutions and political history for example. No national citizenship "scores" on one civics test are contemplated. Instead a series of national profiles will be produced in which one nation may do well on one subtest, poorly on another, and just average on another. This will produce more objective international results which will be useful nationally and internationally.

A selection of test questions has been made from the draft pretests to illustrate the types of questions which will be used to sample content in the five major areas of the civic education inquiry. These are attached in Appendix H-2. The pre-tests are presented in Volume II, Booklet 5.

Three pretest forms of 35 items each were prepared for population 1, three forms of 45 items each for population II, and three forms of 50 items each for population IVG (General), and five forms of 50 items each for population IVS (Specialists). Each of these 14 different forms have seven items in common (which sample the content and abilities dimensions of the test specifications) so that some estimate of the average ability level of the groups being tested can be made both horizontally and vertically, i.e., within the same age groups and among different age groups which are being exposed to different forms of the pretests. All pretests were designed to be completed within a class period of 45-50 minutes duration. The time taken to complete the test as well as the average difficulty scores will also give us some indication of the relative difficulty of the various test forms.

7.5 Criteria for Selection of Items for Final Draft Tests

It is necessary to describe the following general criteria which will be used in selection of items for the final drafts of the civic education cognitive tests for populations I, II, IV, and IVS:

1. Is this item measuring what we say it should be measuring, i.e., is it valid?

2. Is it important to know this particular piece of information, i.e., on the basis of committee comments, national center documents, reviewers' comments, etc.?

3. Is the item working well cross nationally, i.e., do the item statistics indicate that the item is discriminating well between high and low scoring students, is the item within an acceptable range of average difficulty in each nation, are the distractors working, etc.

These and other such questions will be asked and answered after the item analyses from all participating countries are available and after the civic education committee has had a chance to review these results. By the spring of 1969 this task should be completed and the final draft tests available for publication and the dry run by that spring.

7.6 The Development of Affective Instruments

As we have seen, the analysis of the content of the country documents showed that all the countries concerned emphasized education for citizenship and the broad ideological area of equality, egalitarianism or democratic values. The International Civics Committee agreed therefore -

I. to regard these two sets of values as a potential common core for cross-national assessment;

II. to carry out pilot interviews in schools which would lead to a more detailed conceptual framework and a better understanding of children's developing thought processes in this area;

III. to proceed somewhat separately from the cognitive domain; while recognising the close interlinking of perception, cognition, values, and behavior tendencies; this was done purely for reasons of convenience and does not imply a misleading dichotomy.

Subsequently, the pilot interviews took place in the UK, Sweden, W. Germany and Italy. (See Appendix H-3). These were followed by a meeting in Apeldoorn in November, 1967, at which a conceptual framework was drawn up and agreed to, and at which a certain amount of item writing took place. Since then, a number of pilot questionnaires have been developed and tried out, in the UK, W. Germany and the USA. The results of the analysis of these data will shortly be forthcoming, and will lead to fewer and shorter measures with known characteristics in several countries.

The pilot interviews showed very vividly that political awareness and perception of the community develop only gradually. The young child tends to focus on himself and the people he knows, and does not perceive the community as an organised whole, its services, economy, education system, and its many competing interest groups. He trusts adults and does not question why or how things are provided, he lacks a sense of historical development and future progress, and he cannot conceive of adult conflict. His cognitive style is such that he tends to personalise and 'concretise', i.e., to see political events in familiar personal terms rather than as broad principles or institutions: if the British government made a mistake, the landlord of the House of Commons would evict them. Everything is seen in simple, stereotyped, good/bad terms at first, with little awareness of finer differentiations or of the possibility of criticism and improvement. Politicians, elections, fellow citizens, his own and other countries are all seen by the child in very simplistic terms; only later does he become aware of broader issues such as individual freedom vs. the rights of others, civil liberties, majority rule, social justice, or the functions of the State.

In addition, therefore, to measures of our two broad areas of concern (citizenship and egalitarianism), we have developed measures which - hopefully - will serve to demonstrate children's growing perceptual awareness of their communities. We, for instance, have asked a number of questions which will allow the child to answer in personal or in community-oriented terms, to show his perception of politicians and of elections, his awareness of historical development, or his feelings of closeness/distance in relation to the national government. Here are some examples:

Q. Why do wars sometimes break out? (one answer: "Because sometimes countries are too greedy" - tendency to personalise.)

Q. Why do births, deaths and marriages have to be officially recorded? (one answer: "So that they can know how many births they had and, if too many, try to control it" - less personalized, more community-oriented.)

Q. Why does each person have only one vote in an election? (one answer: "Because if they had more than one vote they would be voting forever" - tendency towards concreteness rather than seeing the abstract principle.)

A set of questions asking "How long have we had each of the following in our country?" - intended to show awareness of progress and historical developments (For instance, about a quarter of a W. German group of Population II pupils answered, "We have always had it" in respect of a King or Queen, women working in factories, our political parties, the jury system in courts, advertisements, hospitals, and newspapers, while 80% said "We have always had voting for men.")

A set of semantic differential rating scales dealing with the child's perception of a. his local town council, b. the national government, and c. his local education authority, covering such aspects as closeness, friendliness, warmth, understanding, efficacy, responsiveness to protest, social justice, power, strictness (and their opposites).

The analysis of pre-tests is now proceeding, with a view to improving and shortening these measures, and turning open-ended questions into closed ones.

The above examples will by implication have brought out another important point: that great care is needed in developing measures of values and precepts in children, because these are subtle processes which are easily biased by instruments which are too highly structured or which impose a pre-conceived, adult framework. For this reason, a variety of techniques has been employed: projective, forced-choice, graphic, etc., and many areas have been covered more than once. Thus we hope in due course to develop the methods which are most suitable for each age-level.

The development of the measures for the two main affective areas (egalitarianism and citizenship has followed similar stages. First, a detailed conceptual framework was drawn up, based on the content analysis of the country documents and on the pilot interviews; next, a variety of instruments was designed using different techniques (projective questions, attitude statements, check lists, forced choice items), which have been tried out in W. Germany, the UK and the USA; finally, as the result of pre-testing and analysis, a shortened and improved set of questionnaires will be finalised for the 'dry run' in 1970.

The detailed conceptual framework -- and hence the measures related to it -- dealing with egalitarianism or democratic values covers principles of equality, civil liberties, majority rule, tolerance of diversity, racial equality, etc. Here are some examples of items in this field:

> A set of alleged quotations was offered, asking children if they might give their support to such a person, or not, and why. For example, "I think you should stay out of politics," said Mr. C, you women have enough to do, looking after the home and bringing up the children." Would you give your support to Mr. C? (One answer: "No. He thinks that women should not vote and everybody over a certain age is allowed to vote and if women want to vote they should be left alone." Acceptance of egalitarian principle.) Another example: "What is the point of giving a good education to boys from poor families?" asked Mr. J, "it will only make them want things they can't ever hope to get." Would you support Mr. J? (One answer: "No. He thinks that boys from poor families should not have an education like rich boys families and I think they should all have the same education" - Acceptance of egalitarian principle.)
>
> A set of attitude statements, e.g.,
> The courts of law should give special treatment to people who are very rich.
> Most people can be trusted to use their vote in a sensible way.
> Every person should have the same chance as every other person to get ahead in the world.
> People of different religions should be allowed to pray in any way they like.

The analysis will show the factor structure of these measures, the degree of inter-correlation between them, and the suitability at different age levels. It is perhaps worth mentioning that in these instruments the emphasis is not on the correctness of the child's knowledge, but on his values and attitudes, on his own developing political ideology.

The conceptual framework regarding citizenship values covered such areas as political participation, willingness to serve the community, obedience to the law, loyalty, the functions of criticism, interest in fellow citizens, interest in other countries, etc. From the pilot interviews we already know that different aspects of the citizen role are stressed to children in different countries, and so we have first of all designed an inventory dealing with the child's percepts of the components and width of the Good Citizen role, in terms of stated ideals. For example, the child is asked to indicate whether any of the following are included in 'what you mean by a good citizen': A good citizen:

obeys the law
is always polite
loves his parents
votes in every election
works hard
has good table manners
minds his own business
is willing to serve on a jury

and so on (34 items).

We stress once more that we do not score these items for 'correctness,' but for width and emphasis: there may well be a set of 'common core' items, which are heavily stressed in all countries, together with variations in the degree of inclusion of some of the remaining items, from country to country and for children of different ages, intelligence levels, backgrounds, etc. In other words, it is a perceptual technique.

Next, we made several attempts to design a suitable instrument to measure the child's own adherence to citizenship values. There are difficulties here, since the child is not yet old enough to take the role of a citizen, and so we have to ask him what he would do, whom he would support, how he would react, and so forth. This inevitably entails elements of unreality, and we certainly do not assume that the instruments have predictive validity for the child's behaviour as a grown-up; all we can hope to find out is what his attitudes are now, in anticipation of his role of citizen. To this end, we designed several instruments employing different techniques (forced-choice, projective, open-ended questions, attitude statements) but covering much the same ground: voting, and political participation in general; efficacy, the value of criticism; standing up for citizens' rights; interest in national and in foreign affairs; resolution of conflicts; law enforcement; loyalty to country; interest in welfare of fellow citizens; etc. Here are some examples:

> "There are millions of other people who will vote in this election," said Mr. A, "so it does not matter whether I vote, or not." Will you think like Mr. A when you grow up? Yes/No/Don't Know.
> Why is that? It is because. . . .
> Participation: voting in elections. Projective technique.)
>
> If there were a general election, would you feel
> (a) that it makes little difference whether you vote, or not?
>
> OR
>
> (b) that you must vote whenever you can?
>
> (Participation: voting in elections. Forced-choice technique.)
>
> Lots of little elections are not important enough to bother about.
> So many people vote in a general election that it would not matter whether I voted or not.
> If a person doesn't care about the result of an election, he should not vote in it.
> It isn't so important to vote if your party is certain to lose.
>
> (Participation: voting in elections. Attitude statement.)
>
> "If you try hard enough," said Mr. H, "you really can make a change in the way the country is run." Will you think like Mr. H when you grow up? Yes/No/Don't Know.
>
> Why is that? It is because. . . .
>
> (Efficacy. Projective technique.)
>
> If something is wrong, it does not help to complain about it to the authorities.
>
> (Efficacy. Attitude statement.)
>
> Our government

pays attention to complaints	:___:___:___:___:___:	doesn't pay attention to complaints
can have its decisions changed by ordinary people	:___:___:___:___:___:	can only have its decisions changed by powerful people

(Efficacy: Semantic Differential rating technique.)

We have also designed a confidence index, in which the child is presented with a number of minor social problem situations (seeing a house on fire, having a noisy neighbour, wanting to meet the local parliamentary representative, borrowing money, getting a marriage licence) which may require him to deal with some aspect of authority or community services; in each case, we not only ask what he would do, but also whether he 'would feel very sure what to do,' 'fairly sure what to do,' or 'uncertain.' As before, we are less concerned with the 'correctness' of the answers than with the child's self-confidence in his ability to operate the system, deal with civil servants, etc.

Finally, we come to the politicization determinants, the aspects of the child's life in and out of school which will need to be measured in order to help account for differences in the dependent variables. There will be special sets of questions in the Teacher Questionnaire but, unfortunately, there can be no Parent Interviews (except, perhaps, in the longitudinal study). We have had to restrict ourselves, therefore, to questions which can be answered by the child himself, as part of the Student Questionnaire. Here we have the following variables:

Classroom practices -- democratic/authoritarian. (A set of 60 True/False items abstracted and adopted from the Pace-Stern index of institutional characteristics, dealing with participation in school decisions, demands for obedience and respect, stress on rules, punishment, free discussion, etc.)

Peergroup practices -- how children cope with settling differences or changing rules among themselves.

Mass-media interests in current affairs, etc. (cinema, radio/television, and reading habits).

Child's interests and activities in current affairs, local and national politics (e.g., help to collect funds, join a protest march, make a speech at a mock-election, collect signatures for a petition, etc.).

Frequency of discussion of current events with friends, with parents, and with teachers.

Rejection of adult values, identification with the Teen-ager role. (40 behaviour items dealing with rebelliousness, precocity and disobedience vs. conformity).

'Peak learning experience' (a set of open-ended questions dealing with striking events, persons or experiences which have helped to influence the child's attitudes to politics and citizenship.)

Efforts are also being made to develop measures of authoritarianism (F), self-esteem, and need achievement.

Attitude research with children is never easy, perhaps least of all when we are dealing with such sensitive areas, and when the results have to be comparable across countries. Comparisons between different age-groups will also be difficult, since attitudinal items suitable to one age-level may not be suitable to the next. Moreover, as we have seen, children develop their political percepts and awareness only gradually, and we must avoid the danger of asking questions concerning issues or attitudes which the child has not yet grasped.

We hope that the instruments which we are developing will produce not merely quantitative scores to make cross-country comparisons possible and meaningful, but also qualitative scores or profiles which will show differences in the way that children see their communities, at different age-levels in different countries, the different emphases they place on certain sub-areas, and so on. In addition, we hope to relate these results to the politicization determinants, to be able to account for some of these qualitative differences.

All this will require careful analysis -- much of it from the typed-out responses rather than from computer output sheets -- and, where necessary, further pre-pilot runs on selected instruments. The final set of instruments which will emerge should be much shorter, and will have known statistical characteristics in several countries.

7.7 Interrelation between Cognitive and Affective Domains:

For the moment, the development of the tests and questionnaires in the Cognitive and in the Affective domains is proceeding separately, though there is a continuous exchange of information, of comments, and of criticism. We are not, however, losing sight of the eventual need to arrive at a more integrated picture of the child's political socialization -- perceptual, cognitive, and affective. Since this is likely to raise considerable methodological and analytic difficulties we shall, in due course, need to mount a small, special pilot study to see how our various sets of scores and results can best be brought together.

This special pilot study cannot take place until the various instruments have been much refined and shortened, so as not to overburden the children. Knowledge and attitudes may develop both separately and together, in ways that are isolated or integrated to varying degrees; we hope that this pilot study will show us how these processes of development in the child can best be assessed and expressed.

7.8 Status of Instruments

Many delays have been experienced with the pre-testing. Final cognitive instruments will be ready by April, 1969 and affective instruments in October, 1969. The instruments given in Volume II, Booklet 5, are the pre-test versions.

CHAPTER 8

Development of Questionnaires and Attitude and Descriptive Scales

8.0 QUESTIONNAIRES

The purpose of the various questionnaires is to collect information on background variables about the individual students being tested, the teachers or the schools involved in the testing program. In a cross-sectional survey as the present one, these variables serve as the independent variables or "inputs" for the analysis. The IEA Council decided that such information would be collected from individual students about their home backgrounds, subjects being studied at present, etc.; from all teachers teaching particular IEA subjects in a school about their training, use of teaching time, etc.; and from the school principals about the school organization, financing, etc. The specific information to be collected, of course, depends on the hypotheses formulated either explicitly or implicitly for the research. The international subject area committees formulated many hypotheses both among their own members and with the aid of national subject area committee members. These hypotheses were subject matter orientated. The IEA Council formed from among its own members a Hypotheses Committee, whose job was to review the literature both in terms of theories of education and in terms of previous research, so as to formulate other hypotheses, most of which transcended subject matters and were to do with the structure of educational systems and various educational practices. At the same time IEA was responsible for the convening of a special conference called the "Cross-National Conference on Education, Manpower and the Economy," where social scientists from different disciplines were brought together to examine the results of the IEA Mathematics Study and to study hypotheses which could be tested in such a large scale survey on the theory of their own discipline. They were asked also to study the variables which they thought should be included in the IEA Study.

The variables (or in some cases general ideas) produced by these three sources (subject matter committees, hypotheses committee, and the cross-national conference) were fed into an IEA Questionnaire Committee, consisting of Professors Husén, Bloom, Wolf and Dr. Postlethwaite. These variables and ideas were reviewed in terms of:

1) translatable into paper and pencil items,
2) suitability and acceptability for a large scale survey of this kind.

After the variables had been screened, those remaining were transformed into questionnaire items. These were checked and reviewed by various IEA members and then assembled into draft questionnaires for pre-testing purposes.

A set of notes was written to accompany each pre-test questionnaire to indicate to National Centers the various problems which arise with the translation of certain items into other languages. Where certain technical terms had been used, these were specifically defined, e.g., comprehensive school, intra-school grouping, etc.

Most of the items were left in open-ended form in the pre-test drafts, so that once the frequency distributions were known, the responses to the items could be cast into a multiple choice format. Since the school questionnaire was particularly long, and it was thought that school principals might object to taking so much time to complete it, it was suggested that twice the number of principals required should be taken and the questionnaire split into two halves such that one-half of the principals were given the first half and the other half the second half.

One concern in the present phase of the IEA Project arises from the proposed inclusion of ten-year-olds as a population for research. While the participating IEA countries have felt that the study of a primary school population afforded great potential benefits, there was serious question raised as to the feasibility of testing these younger students. Particular concern was raised about the ability of ten-year-olds to provide accurate responses to questionnaire items. Consequently, it was decided that a pilot study be undertaken to determine whether ten-year-olds could provide accurate responses to such items. The design of the study was simple. Several classes of ten-year-olds in five countries were administered a short background questionnaire. Following this, an interview was conducted with the mother of each student tested or a questionnaire was transmitted to the home to be completed by the mother. In either case, the same questions that were put to the student were put to his or her mother. Both parent and child responses were coded and punched onto Hollerith cards and sent to a central processing site. The countries which participated in this pilot study were: Australia, Sweden, Iran, Italy and Finland. In each country, an effort was made to sample ten-year-olds from upper, middle and lower socioeconomic levels. Also, roughly equal numbers of boys and girls were tested in each country.

The results of this pilot study were fairly clear cut. Those questions which asked the student to describe some aspect of his present life situation showed a very high degree of agreement with the responses given by his mother. Thus, the proportion of agreement between parent and child about the father's present occupation ranged from 70% in Sweden to 88% in Iran. The median was 84%. Similarly, the proportion of agreement on the place where the child studied in the home ranged from 62% in Iran to 80% in Sweden with a median of 67%. The median correlation between the child's age as reported by himself and by his mother was +.70. Thus, there was substantial agreement between mother and child on variables related to the present life situation of the child.

There was considerably less agreement between the mother and child on items which were retrospective or prospective in nature. For example, the median correlation between mother and child on the amount of additional schooling desired for the child was +.35 while the correlation for length of attendance in a nursery school was +.44 for those that had attended. However, the proportion of agreement between mother and child as to whether the child had actually attended a nursery school ranged from 83% in Australia to 95% in Finland with a median of 91%.

The fact that mother and child agree in their responses to certain items about the home situation is not conclusive proof of the truth of the response. It is, however, a reassuring indicator. On the basis of the results of the pilot study it was decided that it was indeed feasible to ask ten-year-olds to complete questionnaires that required them to furnish information about their present life situation both in school and at home. However, the number of items in the questionnaires that asked the student to give information about the past or the future should be kept to a minimum.

The pre-test questionnaires were administered between February and July 1968 and the frequency distributions were sent back to IEA, together with a report on the pre-testing of the questionnaires and comments on the particular items. It would take too much paper and be of little interest to show the frequency distributions, even of one questionnaire in one country, but it is of interest to note how countries carried out the pre-testing of the various questionnaires. The following are selected excerpts from the report of the pre-testing of student questionnaires in Chile:

8.1 Example of Pre-Testing in Chile

The sample for this population included only students who are at present in the 5th grade of their basic education and having an age corresponding to the definition of I.E.A. (coded 120 to 132).

For the General Section, Science and Reading Comprehension Questionnaires, students from four schools were tested. All the schools were urban and of Santiago. We only tested the 5th grade, because in this type of school, most 10 years old belong to this level.

Socio-economic status was also considered important and we included students representing different socio-economic standards in this proportion: 43.3% belonged to a low socio-economic level and 56.7% were of middle socio-economic standard.

As for sex, 31.2% of the group tested were girls and 68.8% were boys.

Population II (14.0 - 14.11)

Included students between 14.0 and 14.11 years of age, at present in the 9th grade (first year of secondary education) all those students whose age did not fit this age definition were discarded from the final analysis.

We included both public and private schools and we tried to reproduce the reality of the socio-economic status of Chilean students at this level. Generally speaking, the socio-economic status of the students of this age is a little higher than that of students belonging to population I.

Sex QUESTIONNAIRES	GENERAL	SCIENCE	READING COMPREHENSION AND LITERATURE
Boys	57.2%	63.3%	49.8%
Girls	42.8%	36.7%	50.2%
TOTAL	100.0%	100.0%	100.0%

Population IV (Students in their pre-university year)

The questionnaires for this population were given to the appropriate students as defined by I.E.A.

The total number of schools tested was five schools, all of them public.

The following table reflects the distribution by sex on the sample:

Sex QUESTIONNAIRE	GENERAL	SCIENCE	LITERATURE & R. COMPREHEN.	FRENCH	ENGLISH
II	68.5%	33.1%	34.0%	39.3%	28.2%
M	31.5%	66.9%	66.0%	60.7%	71.8%
TOTAL	100.0%	100.0%	100.0%	100.0%	100.0%

The translation of the Questionnaires was done by three teachers of English. The revision of each translation was performed by a Committee - generally speaking - formed by the same persons working at the qualitative analyses of the tests. The administration of the Questionnaires for populations I,II and IV was left in hands of the teachers of classes tested. For population I, the General teacher undertook the administration of the three Questionnaires. For populations II and IV, the teachers of each area involved administered the questionnaires.

All the students registered their answers in the booklets.

8.2 Reviews of Pre-Testing and Use of Unscaled Variables

The responses to each question from each country were collated at the IEA Center onto large master sheets. At the National Technical Officer meeting held in September 1968 the questionnaires were also reviewed and it was at this point that it was decided that for certain variables it was impossible to create one international coding scheme into which all practices could be categorized. It was decided, therefore, to designate such variables as unscaled variables. In this case the National Center was asked to formulate in national terms the question which would obtain the necessary information for the variable. At the same time the National Center had to develop a classification scheme for the coding of the responses. The number of categories allowed was given by the International Center. Wherever an unscaled variable was to be used, the National Center had, of course, to return the classification scheme to the IEA Central staff. An example of an unscaled variable is father's occupation. This did not work well in the Mathematics Study and it is felt that the system of unscaled variables will allow greater flexibility and at the same time more meaningfully interpretable data. The question of how to deal with such variables in the analysis arises, since the categories used in different countries will be different, but on the assumption that a single Procrustean mould would be likely to mask their effects in the analysis, it is agreed that it will be better to use the categories as they actually exist in each country. Thus, suppose that in country M there are four categories of teacher education, A, B, C and D. In the i^{th} school in that country they will be represented in the proportions p_A, p_B, p_C, p_D. D can be taken as standard, and p_A, p_B, and p_C put in the regression betweeen schools, to obtain the corresponding terms in the regression equation:

$$Y/s_Y = \ldots\ldots\ldots\ldots b_A\ p_A/s_A + b_B\ p_B/s_B + b_C\ p_C/s_C \ldots\ldots\ldots\ldots$$

in standard form or alternatively

$$Y = \ldots\ldots\ldots\ldots (b_A\ \ s_Y/s_A)\ p_A + (b_B\ \ s_Y/s_B)\ p_B + (b_C\ \ s_Y/s_C)\ p_C$$

in metric form.

Furthermore, provided A, B and C are kept together as a group in the regression, and are not allowed to be separated by other variables through the use of descending partials, the addition to R^2 made by the group can be obtained when it is in any position, including the first and the last, with respect to other groups of predictors.

Consequently comparing the effects of teacher training in different countries can give

Country	M	N	O	etc.
	b_A	b_E	b_J	
	b_B	b_F	b_K	etc.
	b_C	b_G	-	
	-	b_H	-	

This is in standard form. Alternatively the same table could be given in metric form.

It is also possible to give

Country	M	N	O	etc.
	k^2_M	k^2_N	k^2_O	etc.

Here k^2 is the addition made to R^2 by the group of teacher training categories when the group is taken first, last, or in any other position in the regression thought fit.

There is no need to have the same number of categories in all countries. The number of regression coefficients will always be one less than the number of categories, but there will always be a single k^2, which will be maximised by taking the categories as they actually exist in each country, instead of trying to force them into a single mould.

The results of the pre-testing were reviewed and the items for the final questionnaire decided upon. All of the responses to items were then closed on the basis of the frequency distributions obtained in the pre-testing, except for questions such as age, grade, father's occupation, and expected occupation, which were to be given by students.

It was felt that ten-year-old students could not be expected to fill in their replies on an answer card and they will fill in their answers on the answer booklet and these will be punched onto punched cards afterwards by the National Centers. The same applies to the school questionnaire, but the reason in this case is that some of the responses would require four columns of answers and therefore it was much simpler for school principals or their deputies to fill in their answers directly on the booklet. In all other cases the answers will be put directly onto the answer cards to be used by IEA.

The examples of the final questionnaires and coding and punching schemes where appropriate, together with accompanying notes, are included in Volume II, Booklet 4.

8.3 Attitude and Descriptive Scales

The construction of attitude scales with international validity presents some very serious methodological problems. Firstly, there is the problem of defining the dimension that the scale is intended to measure in terms which are equally meaningful in the different cultures for which use of the scale is envisaged.

There is a problem inherent in the translation of attitude statements from one language to another. Although it is usually possible to make a fairly exact translation of the denotative meaning of a statement, it is often difficult to do this without considerably increasing the nature and complexity of the statement from a linguistic point of view. It is even more difficult to also translate the connotative implications, so important in attitudinal work, from one language to another.

Once an adequate scale has been drafted, it is necessary to validate it within each of several different countries. Unfortunately the criteria for validation (even content validation) may appear very different to the recognized local experts in each culture. These views must be reconciled and a compromise achieved.

Finally, there is the problem of "acceptability." Many Western European countries and others are much less used to attitude scales than is the United States and the use of scales which seem quite innocuous in the United States may be vetoed in other countries by the researchers themselves, the Ministries of Education, the teachers' unions or even the school principals.

Despite these very great difficulties, it was decided that I.E.A. should persevere with the development of suitable international scales. The "replications-over-countries" structure of the I.E.A. project meant that even if a particular scale was administered in only a few countries, the results of this could still be valuable. I.E.A. analyses are by and large carried out on a country-by-country bais.

It has been suggested that although we seek to measure the same dimension in different countries, it would be possible to do this by using different sets of items. It seems clear that some items are more appropriate in some countries than in others. A possible procedure is:

a) Carry out factor analyses within each country
b) Match factors across countries to identify the one that is of interest
c) Choose for each country the set of items that seems most adequately to measure this factor
d) Establish norms for these different scales
e) Administer the different scales in different countries and standardize the results

This work on the scale construction has been carried out during the period from September, 1967, until the present. Unfortunately, the need to produce scales by the end of Stage 1 (December, 1968) together with the considerable variation of school schedules (and therefore of possible testing occasions) in the different countries, made it impossible to carry out the full series of steps outlined above in all cases. The basic shortage of time was aggravated in some countries by school strikes, committees of revolutionary students, unexplained postal losses, and the like. We have had, therefore, to take certain shortcuts particularly with regard to Step 2, but it is emphasized that no scale will be considered to be in its final form until it has been pre-tested in all the relevant countries.

The statistical analyses of the pre-test data for these scales have all been performed by the I.E.A. data processing unit which has been stationed in London during this period. The analysis that has been performed is of two different types: a) a factor analysis to investigate the dimensionality of the attitudinal space concerned, and b) a standard item analysis giving response distributions for each variable together with internal consistency figures.

8.4 General Attitude Scales

These scales are intended to be administered to all students involved in the I.E.A. study:

1. Like/Dislike School Scale. After an initial trial of an open-ended instrument in England, a pool of some thirty statements was formed. These statements were put into two forms: multiple choice of the Likert type and true/false. Further piloting in England suggested that the true/false items were working better and ten of these were chosen for full pre-testing. Results so far analyzed from seven countries indicate that the scale is reasonably satisfactory. Reliability estimates had a median value of .80. A slight rewording was proposed for one of the items and two other items proposed by National Centers have since been added so that the present form of the scale contains twelve items. In this form a scale will be administered during the dry-run.

2. F-Scale. The desirability of including such a scale in the battery was discussed during the spring of 1967. Some piloting work was carried out during 1968 in England using items taken from the work of Adorno, Franl-Brunswik and others. A draft scale for international use was sent to several National Centers for administration and to all National Centers for comment. There was a great deal of criticism of this particular scale and indeed of the suggestion that such items be used at all in research projects like I.E.A. As a result of this, the matter was discussed by the I.E.A. Council at its meeting in Hamburg during December, 1968. At this meeting it was resolved, because of the good-will that I.E.A. might lose by administering this type of scale in many of the participating countries, that further work on the F-scale be stopped. Thus, no F-scale will appear in the final I.E.A. battery.

The problem with this arises in step b). Factor matching is difficult to accomplish, and in the circumstances of I.E.A. there would be very little evidence on which the matching could be based. However, a firm base for matching would be important because I.E.A. would clearly be open to attack when the results were reported, and it was clear that different items were used in different countries. It seems that we would not be justified in adopting a policy of varying the items between countries unless we had very sound substantive or statistical evidence proving the equivalence of the scales so formed.

For these reasons the I.E.A. decided to concentrate on the development of scales which were directly translatable into the different languages and cultures involved in the project. However, a degree of flexibility was left to cope with local translation difficulties. Although a single international form for each item (together with translation units) was produced, individual National Centers were allowed to vary the wording of particular items if local circumstances so required. It was not expected that any National Center would modify more than about 5% of the items in this way, and these slight local variations were thought unlikely to affect the general international validity of the scales.

A set of procedures was developed for the construction of such scales. This comprised three separate steps:

Step 1. The development of a multiple-choice item pool from data gathered directly from students. Frequently, this material would come out of interviews or from a questionnaire of the sentence completion type. In most cases this work was performed within one country. When a suitably large pool of items was developed, they were given a preliminary trial with students of the appropriate age but still within the one country.

Step 2. From these results a draft scale of perhaps twenty items was produced for pilot testing in two or three I.E.A. countries with different languages and cultures than the country of origin.

Step 3. The evaluation of the statistical performance of the items together with the detailed comments gathered from the National Centers who had conducted the pilot testing led to a revised scale which only then was put forward for pre-testing in all I.E.A. countries. It was expected that this pre-testing would uncover minor problems of translation in cartain countries but would not lead to more than local revision of a few items.

4. Self-esteem Scale. Pilot work was carried out in England with both a true/false and a multiple choice form, each developed from a pool of attitude statements about the self. Of these, the true/false scale seemed to be superior but the results were in general very unsatisfactory. After discussion by the I.E.A. Council in Hamburg, it was decided that as this scale was not regarded as crucial for the work of the Association, it should be dropped from the test battery.

8.5 Science Attitude Scales

The International Science Committee proposed two types of attitude scale: one, a behavioral attempt to approach "Interest in Science." Pilot work carried out in England in collaboration with an "Attitudes to Science"project of the National Foundation for Educational Research in England and Wales suggested that two quite separate dimensions of attitude could be discerned within the latter. In simple terms there were an abstract view of science as one of the important agents for change in the world, and the more personal view of science as a subject for study in school. Although the testing of these dimensions was originally tried out with students from Populations I, II and IV, it seemed clear that they were not working satisfactorily with Population I (who, for the most part, had not had any formal study of science in school). For this reason, the final forms of these two attitude scales are intended for use with students aged 14 and upwards.

1. Science in the World. In this scale science is seen as a force for good or evil in rather abstract terms. The items that formed the draft scale for pre-testing were chosen following an English factor analysis on a pool of some 300 attitude statements. A six item scale was tried out in twelve of the I.E.A. countries and the results seemed very satisfactory. The twenty-four reliability estimates for Populations II and IV had a median value of 0.68 which seems to be adequate for a survey such as this. The median of the eight estimates tested are proposed for use in Stage 2 without further modification.

2. Science and the Self. This scale deals with the personal appreciation of science as a subject for study in the school. The development of this scale followed the procedures outlined for "Science in the World." The reliability estimates which emerged from the pre-testing had a median value of 0.71. Once again, the scale is being retained in its pre-testing form for its administration in Stage 2.

3. Interest in Science. This scale is intended for use with all populations. The International Science Committee produced a lengthy list of scientific and science-related activities which seemed appropriate for children. To this were added questions concerning school performance in science subjects as compared with other subjects, and questions designed to investigate the extent of the reading of scientific materials outside the classroom. After pilot testing in Finland, Japan, and the United States, the original pool of some sixty items was reduced to one of twenty-four. The revised scale was then circulated to all National Centers for pre-testing. The results for the ten countries which have so far been analyzed suggest that apart from a few minor changes in wording in certain countries the scale is quite satisfactory. The reliability estimates with Population I are about 0.7 and with Populations II and IV about 0.8 and above.

8.6 Literature Attitude Scales

The scales were proposed by the International Committee on Literature. Their development is discussed below. It must be remembered that the Literature Committee did not concern itself with tests or scales for Population I.

1. Transfer of Affect. This scale seeks to measure the extent to which students see themselves as participating in the fictional literary situations they encounter, and the extent to which fictional characters are imagined to play a real part in the students' lives. We are concerned, therefore, with the amount of transference which takes place between the student and the literary word he confronts. We are interested in the extent to which he interprets reality in terms of his literary experiences. Twenty-one items were prepared by members of the committee working in the United States. These were given a field trial in Belgium and Iran and, as a result, a revised scale of ten items was tried out in four countries: England, Sweden, Iran and Finland. The median reliability for a scale comprising these ten items was estimated at 0.63. The Literature Committee, after discussing the results, proposed minor alterations to one of the items and after making this change, the items were presented for use in the dry-run.

2. Censorship. This was a scale proposed by the committee to investigate the student's attitudes toward censorship in its various forms. Twenty-nine items were drafted by committee members in the United States and were circulated to National Centers for pilot testing. It unacceptable and, elsewhere, serious doubts were expressed as to whether such an instrument could be used in the schools. In view of these criticisms, the committee reconsidered the matter and decided to withdraw the scale. No further work has been done.

3. Interest in Literature. The scale seeks to measure the extent to which the students read books for pleasure and engage in other literature--related activities. The twenty items originally drafted in the United States were piloted in Belgium and Iran. As a result of this trial the scale was reduced to thirteen items with substantial rewording in some places. Pilot testing was then carried out in England, Sweden, Iran and Finland with moderately satisfactory results. The median reliability was 0.57. Various suggestions were made to improve the structure and wording of some questions in an effort to improve the reliability. These changes have been made in the form of the scale proposed for the dry-run.

8.7 French Attitude Scales

The International Committee on the Study of French as a Foreign Language requested that work commence on three attitude scales. They were designed to help answer questions about the way in which a student's personality structure could influence the learning of French, and, conversely, how his study of French could affect his general attitudes towards French people and culture. A first scale, Ethnocentricism, was composed of items taken from the work of Adorno and others. The wording had been changed in some places to make the items more appropriate for a cross-cultural study. A second scale was "Attitudes Towards the French People" and these items came in part from the work of W.E. Lambert and others in Canada. The third scale, "Preference for Own Country Over France" owed much to Canadian work and was supplemented by items proposed by the International Committee. A draft instrument of 29 attitudinal statements was formed and this was tried out in Sweden and Scotland. The results were generally rather poor with the reliabilities for subscales being of the order of 0.5. This was disappointing since use of many of these items in Canada and the United States suggested that reliabilities of 0.80 and above should be attainable.

A new revised instrument was composed in which the attitude "Preference for Own Country Over France" was excluded. This was done because of the apparently high correlation between this scale and the "Attitudes Towards the French People." The remaining items were considerably revised and the resulting instrument (6 "Ethnocentricism" items and 12 "Attitudes Towards the French People" items) was circulated for pre-testing in all National Centers. The results from four countries have now been analyzed and the findings are generally satisfactory (median reliability for Ethnocentricism; median reliability for "Attitude Towards the French People."

Concurrently with this work, it was suggested that the Semantic Differential technique be used to investigate more directly the national stereotypes involved. Accordingly, National Centers were asked to pretest an instrument in which "Self," a "Typical French Person," and a "Typical Person of my Own Country" were analyzed. These three were rated on a set of 12 adjective pairs. Analysis of this data and the continuing development of other scales is now proceeding. The decision concerning whether all scales will be used in the French testing has not yet been made.

8.8 English as a Foreign Language Attitude Scales

It was at first thought possible to construct scales which were in general similar to those requested by the French Committee. A complication, however, was that English-speaking people comprise several different cultural groups and it was considered probable that in some countries attitudes towards British people would be very different to attitudes toward Americans. A draft instrument for testing "Ethnocentricism" and the "Attitude Towards English-Speaking People" was pilot tested in Italy and in West Germany. The Italians produced two forms which separately measured attitudes towards British people and attitudes towards American people while the Germans kept the draft instrument in its original undifferentiated form. Results from these pilot testings were very disappointing and strong criticism was received from the Italian National Center since what they regarded as pseudo-political questions were being asked. The English Committee and the I.E.A. Bureau when discussing this matter concluded that the attitude scales should not become involved in political issues and, in consequence, decided to drop scales which contained references to English-speaking people. After consultation between the data processing personnel in London and the chairman of the English as a Foreign Language Committee, a new pool of 47 attitudinal statements about English was constructed. These deal specifically with the study of English in schools and with the English language as a cultural entity. This draft instrument was sent to all participating National Centers and the data which is now being returned will be subjected to a factor analysis before a final form of scales can be proposed. This scale will be a National Option. Scales of interest in English as a foreign language and of attitude towards learning English as a foreign language in school are also being prepared.

8.9 Civics Education Attitudes

These will not be dealt with here and readers are referred to Chapter 7

8.10 Descriptive Scales

The purpose of the descriptive scales is quite separate from that of the attitudinal scales described above. These are intended as measures of the school environment. It has been found (by Pace and Stern as well as by previous I.E.A. studies) that there are some aspects of the school environment for which one can get better data from the students themselves than one could hope to get from the principal. We can ask the principal to report on the enrollment of his school and on the academic qualifications of his teaching staff but we cannot get a reliable assessment from him as to the authoritarian or permissive atmosphere that is characteristic of his school. However, if we collect data from a group of students in school and then average it, we come up with a measure of at least moderate reliability and validity for the school as a whole. The instruments themselves are similar in appearance to conventional attitude scales and are presented to the students as such.

The criteria for success of these scales are that they should discriminate between schools and be consistent within schools. This can be expressed most simply as a F-ratio of between to within school variance. In the development phase of the scales, F-statistics are worked out for each item separately within each population and each country, and in general, items which have consistently high F values (say 3 or above) are retained. Similarly, F-ratios calculated for the total score on a scale gives a measure of the general effectiveness of the scale and for the I.E.A. work, a target of F is = 10 was set for each scale.

8.11 General Descriptive Scales

Two environmental measures were proposed. One, an "Authoritarian-Permissive" scale and the other on a "Subject Oriented-Student Oriented" dimension. After pilot testing in England, an instrument of items covering these two scales was pre-tested by National Centers and the results were subjected to Analysis of Variance, Factor Analysis and a standard Item Analysis. It became clear that the dimension "Subject Oriented-Student Oriented" was meaningful only in England and Scotland in the sense that these were the only countries in which there appeared to be real distinctions between schools. The scale dealt with characteristics such as: (a) the students change classrooms to meet a particular teacher rather than vice versa, and (b) the availability in the classroom of some non-standard subjects for students with particular interests and needs. These needs appear to be more a consequence of national education policy than of the particular school regime. The chief problem with the authoritarian measure was that practices such as "students standing up when a teacher entered a room," which proved highly successful in discriminating between schools in some countries, were regarded as socially dangerous in other countries where the practice of standing for a teacher was universal. "The pupils' standing up has nothing to do with authoritarianism because any other behavior would be extremely bad-mannered." It was apparent, therefore, that the items for this scale must be chosen with great care and that there might have to be local variations in order to have effective items in every country. A revised scale of 12 items has been proposed for use in the dry-run and, using the pre-test data as a guide, it is estimated that this scale should achieve F-values of between 8 and 10 on average.

8.12 Science Descriptive Scales

Two descriptive scales were proposed by the International Science Committee. The first covered the classroom teaching with the textbook or lecture at one end of the continuum and an experimental approach using demonstrations, laboratory work, field work, etc., at the other. This is a fairly straightforward scale of 12 items which achieved a median F-value of 8.3 during pre-testing.

The second scale which deals with the nature of laboratory work is applicable only in those schools which have science laboratories. At one end of the continuum is a laboratory session in which the students are given detailed written instructions which cover all the work that they are allowed to do and at the other end is a laboratory in which the students design and carry out their own experiments without intervention from the teacher. This scale has 8 items. It appeared to be satisfactory in pre-testing and has come forward without modification for use in the dry-run. The attitude and descriptive scales are presented in Volume II, Booklets 4 and 5.

CHAPTER 9

Hypotheses, Sampling Designs and Between Country Analyses

9.0 Development of Hypotheses

The IEA Project can be conceived of within the general framework of an empirical approach in comparative education. Until recently comparative education confined itself to qualitative descriptions of various educational systems and to analyses which then were limited by the available data. Comparative education should, of course, try to achieve what has long since been accomplished in other social sciences, namely to adopt the logico-deductive way of reasoning and the empirical testing of the hypotheses advanced.

In comparative education a set of systems of education, as represented by a group of countries, can be regarded as a large "laboratory" where a wide range of structural and instructional practices are employed. By relating the outcomes of the systems, in terms of specific knowledge, skills, attitudes, etc., to different groups of input factors, such as economic standards and cultural level, teacher competence, equality of opportunity, etc., countries can learn from each other. There are at least three basic prerequisites for such studies. Evaluation instruments applicable at the international level will have to be developed. Representative samples of students at certain age or grade levels will have to be drawn in all the participating countries. The instruments will have to be administered and scores uniformly in all the countries.

In the present phase of the IEA Project hypotheses have been generated in two different ways. The international subject matter committees have been asked to list hypotheses which are closely related to their particular subject fields. Thus, for instance, the reading comprehension committee listed a set of predictor variables which were hypothesized to account for differences in reading performance both within and between educational systems. The list of suggested questions and hypotheses by the subject matter committees are indicated in Appendix I.

The most important work, however, in trying to advance hypotheses thought to be independent of subject fields was carried out by a Hypothesis Committee.which attempted to draw its hypotheses from previous work in the field, from its previous conceptual framework (Husen, 1967) and from the suggestions of consultants researchers in the different National Centers.

On the basis of the experiences gained during the mathematics phase of the IEA Project, it was suggested that a more powerful conceptual framework for the study should be developed. The mathematics phase was designed entirely by educational researchers which tended to overemphasize school and teacher variables in the list of predictor variables which were drawn up and turned into instruments used to test the hypothesis. The educational system must be regarded as a sub-system of the social system at large. In order to develop an input-output model, which hopefully could help in the development of fruitful hypotheses, sociologists, economists and political scientists convened with representatives from the IEA group on two occasions in Hamburg in 1966, and at Lake Mohonk, N.Y. in 1967 respectively in order to suggest a better theoretical model for the IEA project and to list variables within the social matrix, which were thought to be relevant to the outcomes of the educational system. Both cognitive and non-cognitive outcomes of these systems could be related to various groups of input variables, such as expenditure per pupil, teacher characteristics, equipment, etc. (Cross-National Conference on Education, Manpower, and the Economy, 1967).

Previous research carried out, for instance, by the Plowden Committee in England and the Coleman group in the United States, seems to indicate that the major portion of inter-individual differences in school attainments are accounted for by non-scholastic factors, such as home background and parental motivation. Since the socio-cultural variability in the present phase of the IEA Project is much greater, hopefully a general answer will be obtained to the question of how close the relationship is between the socio-cultural matrix and the "productivity" of the educational system. An answer to this question has important implications for educational policy, particularly in the developing countries, because if education cannot achieve considerably more than the ceiling set by the socio-cultural level of a particular country, the place of education in the list of priorities would have to be revised.

It should be pointed out that the second phase of the Project offers a unique opportunity to test certain general hypotheses on the relationship between certain input factors and the outcome of the system, not only across countries representing a wide variety of socio-cultural patterns, but also across subject areas.

In addition to the specific subject matter hypotheses and questions, (See Appendix I), the hypotheses committee formulated hypotheses which transcended a particular subject. In doing this, it stated the hypotheses, indicated which subjects were relevant, the populations for which it should be tested (1,2,3,4,), and whether the analysis should be between students within schools (ST), between schools (SC), or between national systems (N).

These hypotheses will need additional work before the details of data processing can be executed. Insofar as possible, composite hypotheses will be formed from the single hypothesis for the purposes of hypothesis testing. Further committee effort will be needed to complete this phase of our work. The specific hypotheses are listed below and are grouped to indicate whether they refer primarily to the Home and Community background, to the Educational Policy of the System, to the School, to the Curriculum and Instruction, the Teacher, or primarily to specific characteristics of the students.

Home and Community Hypotheses

1. Achievement varies with parental socio-economic status, least in the case of the simple skills and content, most in the case of the complex.
(All subjects) (1,2,3,4) (ST, SC)

2. There will be systematic socio-economic differences between students in 14-year old populations and those in pre-University populations.
(Not applicable) (2 and 4) (N)

3. Achievement of urban pupils will exceed that of rural pupils.
(All subjects) (1,2,3,4) (ST, SC)

4. Achievement, interestes attitudes will be positively correlated with social status of pupils' houses.
(All subjects) (1,2,3,4) (ST, SC, N)

5. Motivation is positively related to the discrepancy between father's occupation and expected (aspired) occupation.
(Science, English) (3 and 4) (ST, N)

6. Pupils whose native language is different from the language of instruction will have lower achievement scores in all subjects than pupils for whom there is no difference.
(All subjects) (1,2,3,4) (ST)

7. Level of achievement in foreign languages is related to family community and school attitudes towards the language studied.
(French, English) (1,2,3,4) (ST, SC, N)

8. Level of achievement in a foreign language is related to opportunities for contact with persons and groups using that language.
(French, English) (1,2,3,4) (ST, SC, N)

Educational Policy of the System Hypotheses

1. The level of achievement at age 14-14.11 is not related to the mean age when compulsory schooling begins.
(Science, English, French, Reading Comprehension) (2) (N)

2. When equal proportions of the age group are compared, countries will not differ in their terminal level of achievement.
(All subjects) (4) (N)

3. Average level of achievement in the terminal group will be lower in countries with larger percentage of age group still in school.
(All subjects) (4) (N)

4. Mean performance on total test in any given subject area of top 5 per cent (10 per cent) of the age group is higher in systems with high retentivity than in systems with low retentivity.
(Science, English, or French) (3 and 4) (SC, N)

5. Sex differences in mean performance in each subject as well as attitudes toward the subject is related to amount of coeducation in a given country.
(All subjects) (2,3,4) (N)

6. The "total yield" (defined as an answer to the question 'How many are brought how far?') is higher in a highly retentive than in a less retentive system.
(Science, English or French) (3 and 4) (N)

School Hypotheses

1. Achievement, interests and attitudes in a school will be related to total enrolment of the school.
(All subjects) (1,2,3,4) (SC)

2. The level of achievement in a given subject is not related to size of class.
(All subjects) (1,2,3,4) (ST)

3. The level of achievement will be higher and the variability will be lower in specialized schools than in comprehensive schools.
(All subjects) (1,2) (SC)

4. School schievement will be positively connected to (a) the proportion of specialist teachers of the subject at school, (b) the provision of laboratories, (c) the provision of books.
(All for (a), science for (b), all for (c)) (1,2,3,4) (SC)

5. Student achievement will be related to per student expenditure for teachers' salaries and for other expenses.
(All subjects) (1,2,4) (SC, N)

6. Achievement, attitudes and interests in school subjects are unrelated to available funds, space or teacher qualifications.
(All subjects) (1,2,3,4) (SC, N)

Curriculum and Instructions Hypotheses

1. Level of achievement will be higher when the number of subjects studied is smaller.
(All subjects) (4) (ST,SC,N)

2. Holding opportunity to learn constant, inquiry-centered methods will produce higher and less variable achievement scores and affective indices than will more traditional ("drill") methods.
(All subjects) (1,2,3,4) (ST, N)

3. The profile of test performance in each country will be related to the national emphasis on each topic (as reported in teachers' ratings) in each school program.
(All subjects) (1,2,3,4) (N)

4. Student achievement in a given subject is (a) not correlated to number of school hours of schooling per week but (b) related to those of a given subject as well as (c) to number of homework spent in the subject.
(All subjects) (1,2,3,4) (ST)

5. Student achievement in partscore (topics and objectives) will be related differently to the number of hours per week devoted to all school homework as well as those devoted to the subject according to the kind of objectives and/or subject-areas.
(All subjects) (1,2,3,4) (ST)

6. Achievement is not related to teaching method or learning method.
(All subjects) (1,2,3,4) (ST,SC,N)

7. Methods of learning involving first hand experience (laboratory, primary sources) are related to interest in subject matter.
(Science, Literature) (2,4) (ST)

8. Student achievement will be higher and less variable when there are provisions for individual differences in learning rates and styles.
(All subjects) (1,2,3,4) (ST,SC)

9. Student achievement will be higher and less variable when the maximum per cent of teachers' time in the classroom is spent in teaching.
(All subjects) (1,2,3,4) (ST, SC, N)

10. There is an optimum amount of instruction in a second language - beyond that the students show little additional progress in achievement, interests and attitudes.
(French, English) (2, 4) (ST)

11. Variations in the achievement profile are related to variations in teachings objectives and methods.
(French, English) (1,2,3,4) (ST, SC, N)

12. In any one subject the achievement, interests and attitudes will be related both to the stated aims of instruction and to the opportunities given for learning.
(All subjects) (1,2,3,4) (SC, N)

Teacher Characteristics Hypotheses

1. Student achievement will be positively correlated to the quality and quantity and recency of training of teachers.
(All subjects) (1,2,3 and 4) (ST, SC)

2. Achievement scores will be higher in schools where teachers feel themselves and have greater freedom in determining what will be taught and how it will be taught.
(All subjects) (1,2,3,4) (SC, N)

3. Students whose teachers have had recent (within 5 years) in-service training in a subject will have more positive attitudes toward and interest in that subject.
(All subjects) (1,2,3,4) (ST, SC, N)

4. Achievement will be related to the amount and quality of teacher education.
(All subjects) (1,2,3,4) (ST, SC, N)

5. Attitude of teachers towards their profession and subject matter will correlate positively to the achievement and motivation of the pupils.
(All subjects) (1,2,3,4) (ST, SC, N)

6. Teachers' attitudes towards (or expectations about) the children's learning are strongly related to achievement in a given subject.
(All subjects) (1,2, and 3) (ST, SC, N)

Student Characteristics Hypotheses

1. Motivation is positively correlated with achievement in any given subject area.
(All subjects) (1 & 2) (ST, SC, N)

2. Total score will be positively correlated with interests and attitudes about the subject.
(All subjects) (1, 2, 3, 4) (ST, SC, N)

3. Students planning to enter vocations in which they consider a subject relevant will make higher achievement scores than other students.
(All subjects except Reading Comprehension) (2, 4) (ST, SC, N)

4. Pupils' profiles of scores are congruent with pupils' cognitive style.
(All subjects) (1,2,3,4) (ST, SC, N)

5. Achievement (also interest/attitudes) is correlated positively with perception of utility of competence in teh subject for upward (social) mobility.
(All subjects) (2,3,4) (ST, SC, N)

6. Achievement, interests and attitudes are positively correlated with estimated or reported importance of the subjects.
(All subjects) (2,4) (ST, SC, N)

7. Students will achieve at higher levels and have more favourable interests and attitudes when they have high esteem for themselves as students than when they have high esteem for themselves as human beings.
(All subjects) (1, 2, 3, 4) (ST)

8. Students will achieve at higher levels and have more favourable interests and attitudes when they have relevant nonscholastic competences.
(All subjects) (1, 2, 3, 4) (ST)

9. Difference between student's occupation expected (desired) and his parents' occupation will be related to the student's cognitive outcome in a given subject.
(All subjects) (1, 2, 3, 4) (ST, N)

10. The difference between student's expected (desired) occupation and his parents' occupation will be related to the student's interests in and attitudes towards a given subject.
(All subjects) (1, 2, 3, 4) (ST)

11. Level of achievement in a foreign language is related to the "linguistic" distance of the mother tongue from that language.
(French, English) (1, 2, 3, 4) (ST, SC, N)

12. Student achievement, interests and attitudes will be differently correlated with the difference between their expected (desired) occupation and their parents' occupation according to the type of school they attend.
(All subjects) (1, 2, 3, 4) (SC)

13. Student perception of his achievement in the given subject will be related to the status and kind of his occupation expected (desired).
(All subjects) (1, 2, 3, 4) (ST, SC)

14. Achievement in any one subject is positively related to participation in non-school ("cultural") activities of the same type.
(All subjects except Reading Comprehension) (1,2,3,4,) (ST, SC, N)

15. The profile of achievement is positively related to the profile of participation in non-school ("cultural") activities of the same type.
(All subjects except Reading Comprehension) (1, 2, 3, 4) (ST, SC, N)

16. Achievement (in any subject) is positively related to student evaluation of intellectual as compared with other possible student roles.
(All subjects) (2, 4) (ST, SC, N)

9.1 Sampling Designs

Much progress has already been made. With three exceptions, set out below, all countries have adopted the general design suggested at the outset in Bulletin No. 3. In this design the sampling proceeds in two stages, the first stage being the selection of schools from a stratified frame, and the second the selection of students from selected schools. For schools the factors of stratification are size, type, sex and region. Each student in the population is given the same chance of entering the sample, by keeping the overall sampling fraction constant. At the same time the design effect is kept small by avoiding wide variation in the number of students selected from each school. This is achieved by using the school measure of size. The probability of selecting the school is made directly proportional to this measure, and the conditional probability of then selecting a student from the school is made inversely proportional. In other words schools are given one, two, three or more tickets in the draw according to their sizes. When a school with one ticket is drawn all its students in the population are included in the sample. A two ticket school has half its students included, a three ticket school a third, and so on. In choosing measures of size the guiding principle is to keep the average number of students selected per school in the neighbourhood of thirty, and to avoid wide departures from this average. On the experience of the mathematics study the prior estimate of the design effect from this plan is in the neighbourhood of four or less, so that a sample of three thousand students from a hundred schools should be equivalent to a single stage sample of 750 students or more, with standard errors in the neighbourhood of 3% of the corresponding standard deviations. This prior estimate will of course be replaced, during the data processing, by a posterior estimate of the design effect obtained by the comparison of half sample and whole sample estimates, the half samples being drawn on the computer according to the full complexity of the sampling design.

The strong point of this design is that the overall sampling fraction is kept constant even when the measures of size are very rough. "Size" means the number of students in the school in the relevant population, and when this is known accurately for all schools "Size" and "Measure of size" can be identical. But where there are no central statistics giving this information an approximation, such as the total number of students divided by the age range of the school, can be used instead, as the measure of size. Roughness in the measure of size will not bias the sample; its effect will be to produce a rather greater variation in the number of students selected from each selected school than would be the case if the measure of size were more accurate. This in turn will produce a rather greater design effect. But even a very rough measure of size will produce a much smaller design effect than no measure.

When the sample of schools has been drawn it will be tested by using it to estimate marker variables, these being variables for which a complete count already exists. An example from Scotland is enclosed to illustrate this process. It will be seen that the sample estimates all accord closely with the population figures, except in the case of very small strata, where the relative sampling variation is bound to be large, and also unimportant.

The samples of schools are being drawn now, well in advance of the date of the field work of the survey, in order to give ample time for checking the correspondence between the design and its execution. If a relatively large number of new schools come into existence between the date of drawing the samples of schools and the date of the field work a supplementary draw will be made to cover this new population.

Since Population 1 occurs in primary schools, and the other populations in secondary schools, separate samples of primary and secondary schools will in any case be needed. Within the secondary field some countries are drawing separate samples of schools to cover each population; others are sampling every population that occurs within a selected school. The latter is the more appropriate procedure in a small country, where the total number of schools is correspondingly small. The main complication that it involves is that a school's measure of size may have to be a compromise between the ideal measures for two different populations. This is not serious, because, as was said above, the measure of size need only be rough, and any reasonable compromise between the alternatives will be good enough. In extreme cases it may be necessary to draw a supplementary sample of schools to cover a small Population IV.

Before the field work begins selected schools will furnish the National Center with an account of the numbers of their students in the relevant populations. In some countries this account will include the names of all such students, and the National Center will nominate the students to be tested, in accordance with the sampling fraction for each school. In other countries, for example in Scotland, this procedure is thought to be too burdensome, and in such cases the selection of students will be made by using their dates of birth. Thus for a school which had two tickets in the draw, so that half its students are to be included in the sample, those would be included whose birthdays occurred on an even (or alternatively of course an odd) day of the month. It is clear that under this procedure there would be very few cases where the number with odd birthdays was markedly different from the number with even birthdays, but when such a case did occur the school would be asked to choose the larger half and discard from it by selecting for discard birthdays spread evenly round the year. Since the student's date of birth is a matter of record, and since students will be required to enter it on a questionnaire, this should be a safe method of selection.

It may be noted that the selection of students from selected schools has been greatly simplified by the decision to postpone the testing of some subjects to a later stage. This has reduced the number of subjects to be tested at the present stage to a point at which it is reasonable to ask all selected students to take all relevant tests. Consequently the elaborate arrangements devised, under the heading of packing and pairing, to avoid placing too great a burden on any student.

It was noted above that three countries have not adopted the general sampling plan without modification. These are the United States, India, and France. Because of the immense size of the country, the United States sampling is to be carried out in three stages. The first stage will be the selection of areas. Within selected areas schools will be selected, and within selected schools, students. The advantage of this plan, in a country so large, is that the more detailed frames for the selection of schools need only be constructed within the selected areas. The design effects will, of course, be somewhat greater than they would be if the same number of schools were selected as the first stage, but this will be offset by selecting a rather larger number of schools.

There is an additional difficulty in the Indian case. Not only is the country of immense size, but the principal language varies from one part to another. Consequently it has been decided to limit the survey to the six states where Hindi is the principal spoken language. Within these states there is already in existence a master sample of communities, and this master sample will be used as the frame for drawing the sample of schools. There is a great deal to be said for both these decisions. To extend the survey to cover other states and languages would strain the resources available to the breaking point. So would a decision to ignore the master sample and construct a new frame from the ground. It is true that the standard errors calculated in processing the data will relate to fluctuations from the master sample, and not from the complete population of the six states, but it seems reasonable to hold that this will not introduce much additional uncertainty, because India, under the guidance of Dr. Mahalonobis whose reputation is international, has been the home of area sampling.

The remaining exception is France. The French scheme follows the ordinary pattern in the first stage, with the selection of schools from a stratified frame using measures of size. But at the second stage the selection within selected schools is made not by selecting students but by selecting classes. This is because the French authorities object to a class being broken up by including some of its members, but not others, in the survey. It should be noted that in France the survey will only cover Population IV, and that the classes with which we are concerned contain the whole of Population IV and nothing but Population IV Furthermore it is possible to grade the classes constituting Population IV within each school. Were it not for this fact the correlations emerging from the survey would be greatly attenuated, since cases will arise when a school with good teachers, good staffing ratio, and excellent facilities is represented by its worst class, and vice versa. But by using the class grade as an unscaled, or categorised, variable in the regression this attenuation can be avoided.

Although the topic does not strictly fall under the head of sampling design a cross reference should perhaps be made at this point to the plans, set out elsewhere, for taking account of the difference, where it exists, between the population in an age group and the population in school. This applies to the other populations as well. It is of course a topic of prime importance, but the plans for dealing with it have been set out elsewhere.

9.2 Analyses Between Countries

The questions we shall seek to answer fall under two heads:

1. Does the evidence support the view that the underlying causal systems are generally similar in all the participating countries? Or do the countries fall into groups, with the countries in each group resembling one another more closely than they resemble countries in other groups? And in either case what are the main features?

2. Is the level of achievement, at each stage studied, much the same in all participating countries, when due allowance is made for the proportion of the population in school, and for the length of schooling, or is there a wide variation between countries, even after these allowances have been made? And how does the new evidence compare with that from our mathematics study, where, for example, the very wide range shown initially between terminal students in different countries was greatly reduced when the allowances indicated were made?

3. There is a prior question which might be regarded as falling under the head of analysis between countries. Are the measures of achievement equally appropriate for all the participating countries? In devising their tests the subject committees have endeavoured to cover all the ground that is thought to be of major importance in any of the participating countries. If they have been successful in this the question of fairness between countries resolves itself into the question of whether the most appropriate weight has been attached to the different parts of

the tests. The part scores have assigned weights, given by the subject committees in their allocation of items to the different parts, and at an early stage in the data processing a component analysis will be carried out to see whether these weights need modification. If so, the modified weights determined by the component analysis will be used to rescore the measures for all subsequent analyses. In the mathematics study no re-scoring was needed, since the correlations between the scores with the assigned weights and those with the modified weights were extremely high.

The answers to the first question above will be sought by comparing the regression analyses for the different countries. From our earlier mathematics survey it seems likely that some of our groups of predictor variables will have effects of roughly the same magnitude, at any rate in those countries which can be thought of as sharing a common culture. Other groups of predictor variables may emerge with notably different effects, in different cultural groups of countries, and yet other groups of variables may turn out to have different effects even within the same cultural group of countries. Sampling theory and past experience alike give grounds for expecting that the estimates for groups of variables will be considerably more stable than those for single variables. In forming the groups of variables we shall be guided in the first place by their nature; for example we shall group together those variables having to do with parents, and those with teachers. In estimating the effect of a group of variables we shall use more than one technique. For example, we shall sum the regression coefficients over each group, and sum the contributions when each variable is taken last in the regression, but we shall also find the sum of the contribution of each group when it is taken in what appears to be its most appropriate place in the order. At the outset there seems to be a fairly strong initial probability that the different techniques will all lead to the same sort of pattern in the grouping of countries. If this expectation is not borne out we shall examine the circumstances in detail, in the hope of obtaining new light on a greatly controverted question, about which there has been a great deal of argument in recent reports and the commentaries thereon. The essence of the controversy is that if a complete incremental account is given of the regression variance the earlier a variable is taken in the regression the more, as a rule, it contributes to the account, so that if an incremental account is given the order must be justified, and it is hard to find arguments that are conclusive for this purpose. The regression coefficients themselves do not depend upon the order, nor, of course, do the contributions when each variable is taken last. But both these modes of statement leave something like half the total of the regression variance upon the joint account of all the variables.

In the mathematics study a common scale was chosen for each variable. In some cases this scale turned out to be inappropriate for some of the participating countries. "Type of teacher training" and "Place of parents residence" are examples. To remedy this in the present study we intend to make use of categorised or unscaled variables, the categories being chosen in each country to suit the actual circumstances. The formal identity between multiple regression theory and the theory of least squares will then be used to combine the estimates from the scaled and the unscaled variables.

In seeking answers to the second main question our prime effort will be to present national mean scores adjusted to take account of differences in the proportion of the population in school at each stage, and of the length of the course to the stage. In the mathematics study we found that these factors, and particularly the first, accounted for more than 60% of the variation in the national means for the terminal students, since school attendance was compulsory, both in principle and in practice, for the other populations in all the countries that took part. But in the present study Chile, India, Iran and Thailand are taking part, and in these countries school attendance is far from universal at any stage. One way of allowing for this fact will be to give percentile scores, the percentiles being reckoned for the whole population, and not merely for that part of the population which is undergoing schooling.

In the mathematics study adjustments by regression and by equivalent percentiles were made, but the table giving the adjusted scores were widely separated from the initial tables of unadjusted means, and therefore attracted less attention. This led to some misunderstanding. We shall avoid this risk in the present study by giving the adjusted and unadjusted tables together, giving the former more prominence and explaining that the unadjusted scores are included merely to show how the adjusted scores were reached. A tentative scheme for the analysis and presentation of the IEA evidence is presented in Appendix J.

CHAPTER 10

PLANS FOR ADMINISTRATION AND DATA COLLECTION

10.0 Introduction

In any survey involving the administration of tests, questionnaires and other measuring instruments, the arrangements for getting these completed by the chosen sample of students and teachers is clearly important. Inevitably there are difficulties in ensuring that the right tests, etc., get to the right students and that all understand exactly what it is they have to do. In surveys that cross country boundaries, especially where many different languages are involved, administrative problems are magnified and great care in planning is necessary if errors are to be avoided.

Fortunately in the planning for the present Phase of IEA, the experience gained in the mathematics Phase could be drawn upon. In that, the administration was considered at three levels and three administrative manuals, one for each level, were prepared. The three levels were (i) the National Centers, that is the National Institutes responsible for the administration of the survey in each participating country, (ii) the Schools, and (iii) the classes or groups of students actually to receive the testing instruments. The administrative procedures employed in the mathematics survey and indeed the actual manuals that had been prepared were invaluable in guiding the arrangements that had to be made for the Phase II survey.

From the outset, however, it was clear that a number of changes in the administration were going to be necessary. Testing instruments in six subject areas instead of one were being prepared and complexities were going to be introduced by the fact that there were more countries involved and that not every country was intending to test in every subject area. Moreover, changes in the methods of obtaining the required information from students and teachers--the use of answer cards instead of answer sheets--meant that different procedures would need to be employed for getting the collected information back to the chosen international computing Center.

The problems associated with the administration of the survey were considered from the beginning of Stage 1. Initially, however, much attention was given to finding an acceptable compromise between having short and perhaps inadequate tests but testing students in every subject and having longer and perhaps more efficient tests but only testing each student in two or at the most three subjects. Ways in which students could be adequately "packed and paired" were devised. That is, procedures were worked out for sub-sampling the students within any given school and arranging them in groups so that the required combinations of subjects could be achieved with no student taking more than the tests in two subject areas. It was, of course, important that the procedures so worked out were administratively feasible and moreover acceptable by the participating schools.

10.1 Separation of Stage 2 and Stage 3

As work progressed on the construction of the cognitive tests in each of the subject areas as well as work on the development of the student questionnaires and attitude scales, it became clear that the original ideas on the administration of the survey would have to undergo some change. The fact that amongst the participating countries some were only wishing to test in one, two or three subjects and others in four, five or six, meant that there were difficulties in trying to arrange a common scheme for pairing subjects. However, it was the realization, as time progressed, that the problems of constructing international tests in some of the subject areas were greater than in others, and that in fact, it was unlikely that the tests in three of the areas would have reached the desired stage of completion by the end of 1968, that finally led to a major change in the proposed plans for the administration.

By May 1968 it was fairly clear that sufficient progress had been made in the construction of the tests for Science, Reading Comprehension and Literature that final versions of them would be available at the latest by November 1968 so that administrative plans could be made for their operational use early in 1969. With the other three subjects, however, - French, English and Civic Education, not all of the construction problems had been solved and the possibility that further pre-testing and analysis work would need to continue to the end of the year, if not into 1969, was apparent. The decision was therefore made to divide the survey stage of the present project into two: Science, Reading Comprehension and Literature would be tested first on one sample of students, and the remaining subjects one year later on a separately drawn sample.

This arrangement had certain additional advantages. In the first place, it meant that the schemes for "packing and pairing" were unnecessary, thus avoiding the very complex administrative arrangements required within schools for matching students and tests. Secondly, this particular grouping of subjects had much to commend it. There was a close association between Reading Comprehension and Literature in that they were both aspects of one subject - Mother Tongue, and furthermore, this subject was in marked contrast with Science, so that, by testing the same students in these subjects some evidence might be gained to examine hypotheses arising from the concept of the "two cultures." Lastly, every country now in IEA had agreed to test in Science--the only subject in which such agreement had been achieved. While a few did not want to test in Reading Comprehension and rather more wished to exclude Literature, this present combination of subjects meant that every country would still participate in the new abbreviated Stage 2, even if some did not continue with the survey program into Stage 3.

10.2 The Need for a Dry Run

In the Mathematics Phase of IEA, 12 countries cooperated. About

130,000 students and 13,000 teachers from over 5,000 schools produced in all about 50,000,000 items of information which had to be processed. The task of ensuring that each item of this information found its way from the student, teacher or school from which it originated to its correct place for processing at the computation center, was no mean task. Even so, it was not accomplished without a large number of mistakes occurring, many of which resulted in a loss of data. A study of these mistakes can lead hopefully to their reduction in the present Phase. Nevertheless, no crystal ball is so clear that all eventualities can be foreseen. Moreover, this present Phase is even more complex administratively than Phase I. Not only are three subjects involved in Stage 2 alone, but there will be nineteen or twenty countries instead of twelve, and moreover, an additional Population (the 10-11 year olds) is to be sampled and tested. At a rough estimate, over 300,000 students will be involved and in all the number of items of information to be collated for processing may increase by a factor of between 5 and 10. This greater complexity will certainly increase the likelihood of mishaps occurring, which either may prevent some information from being collected where it was intended that it should, or once collected, may prevent its transference from its place of origin to the computation center. The aim of careful planning in the administration is, of course, to see that these mishaps do not occur, but when new ventures are made into unknown territory, this is not always possible.

It is for this reason that the decision has been taken to add one whole year into the project in order to try out all administrative procedures. It is intended in 1969 to conduct a "dry-run" for Stage 2. In this, all procedures, including the administration of the survey instruments will be executed in each participating country, except that only small "judgement" samples of students will be used instead of the full-scale probability samples which will be used in the main survey one year later. By having this rehearsal it is hoped to uncover all or most of the snags and problems which are likely to be met, hence enabling a solution to be found before the main survey is undertaken.

Since school years begin at different times in the different countries, it is inevitable that the IEA survey testing will not be carried out everywhere at the same time of the year. Indeed it is clear that testing dates will range from January to November. Since this same range will also apply to the Dry-Run, it was necessary to complete the administrative plans well before the end of 1968 so that those countries wishing to carry out their Dry-Run early in 1969 would not be delayed. Accordingly a document setting out the proposed administrative arrangements was drawn up in May 1968 and sent out to all National Centers for their initial comment. Once these had been received, the Technical Director and Coordinator were able to produce first drafts of the Manuals which were to contain full details of the administration.

10.3 The Three Administrative Manuals

Using the experience gained in Phase I, it was decided to employ again the three administrative levels, and to put the details of the administration relevant to these levels into appropriate manuals. Furthermore, it was agreed that each National Center should appoint a National Technical Officer who was to have the general responsibility for carrying out the survey in his country, and he would have the task of interpreting the meaning of the administrative details included in Manual 1. In addition, a person who was to be referred to as the School Coordinator would have the responsibility for overseeing the administration in each school, and the details of his task would be included in Manual 2. It was foreseen that in many cases the School Coordinator would be the Principal of the school, and in others a single person might act as School Coordinator for more than one school. Finally, at the end of the chain, Manual 3 was prepared for the Test Administrator - the person within the school who would have the final responsibility for actually administering the tests, questionnaires, etc.

Manual 1

This gives the general details of the administrative procedure it was expected that most National Centers would follow. (See Vol. II, Booklet 1). There is, of course, a built-in flexibility to the IEA administration. While it expected for example that all participating countries will administer the agreed tests and other instruments, allowance is made for countries to ask certain questions in the way that suits them best, and for certain "national options" which they can give or not as they choose. This same flexibility governs the methods used to collect and collate information within each country, although for the purposes of Manual 1 certain assumptions are made. It is assumed, for example, that Answer Cards will be used for most student responses - if they are not and a particular country intends to use an alternative method, it must work out the details of the alternative procedure for itself, provided always that it can get the necessary information back to the IEA processing Center in the required form.

Most of what is included in Manual 1 will apply to the main testing in 1970, although it will be used initially in the Dry-Run in 1969. Certain parts, dealing, for example, with the sampling of schools and the printing of questionnaires will have to be modified before 1970 - unless, that is, the Dry-Run itself reveals the necessity for making changes in the administrative arrangements.

Manual 2

Since this is written for a person who may not previously have been aware of IEA, it contains an introduction which briefly explains the aims and purposes of the survey. Otherwise it details the responsibilities of the School Coordinator, and as such the example given in Volume II, Booklet 1 may well need modification if particular National Centers intend to employ alternative arrangements.

It was suggested that in order to ensure that the contents of Manual 2 are properly understood, meetings of School Coordinators are held at which more detailed explanations can be given if required.

Manual 3

As this contains the specific instructions for the administration of the particular tests, questionnaires and attitude scales which are being administered in each country, there will, of course, be variations according to what subjects are being tested. Other variations will occur from the example given in Volume II, Booklet 1 if, in a particular country the testing is being spread over a different period of time, or if additional tests or other instruments are being included as "national options." It is suggested that School Coordinators hold special briefing meetings for the actual test administrators who it is expected in most cases will be teachers.

10.4 The Development of the Administrative Procedures

While it is possible to set down details of the required administration in writing, the complete understanding of what is intended can only be achieved by face to face discussion. Accordingly in September 1968 a meeting was held of the National Technical Officers from all participating countries. Group and individual sessions were held at which all aspects of the administration were discussed, both from the general point of view of the efficiency of the survey as a whole, and specifically as they were affected by the circumstances pertaining locally in the different countries. A number of the main procedures were modified as a result of this meeting, and subsequently further drafts of Manuals 1 and 2 were prepared. These in turn, together with Manual 3, were circulated to National Centers before the final Council meeting of Stage 1 held at the beginning of December.

10.5 GENERAL PLANS FOR CALENDAR AND WRITING ASSIGNMENTS

As mentioned in Section 1, the four years from January 1969 to December 1972 have been split into two stages - Stage 2 and Stage 3. Stage 2 will deal with Science, Reading Comprehension and Literature. Stage 3 consists of English as a foreign language, French as a foreign language and Civic Education. The timetables are as follows:-

	Stage 2	Stage 3
1969: Jan. - Nov.	Dry run	Finalisation of instruments and preparation for Dry Run
December	National Technical Officers Meeting	
1970: Jan. - Nov.	Full Testing	Dry Run
June - Dec.	Analyses	
1971	Analyses and Write-up	Full Testing
1972	Continuation of Analyses and Write-up	Analyses and Write-up
1973	Publication of the Results	

The dry run is spread across a nine month period. Since the final testing must be carried out as near as possible towards the end of the school year, the dry run in most cases will take place twelve months before this, as it is important for it to be an exact replica, only on a small scale, of the full testing. School years end at different times in different countries and in different hemispheres. While the dry run is being carried out in the participating countries, the writing and testing of the computer programs will be undertaken centrally. The computer programs will be written for Stage 2 in such a way that they can be used for Stage 3 with very little alteration. While the dry run for Stage 2 is underway it is hoped to finalise all instruments for Stage 3. It is expected that the instruments for French as a foreign language will be finalised by March 1969, but it may be that some further small testing will be required in the other two subject areas, especially in the attitudinal domains. The results of the dry run for Stage 2 should show us the mistakes which have occurred. Procedures will be developed for putting right the mistakes and therefore in December 1969 or January 1970 it is envisaged that we shall have a meeting of National Technical Officers in order to go through the new procedures so that every National Center is aware of how to correct the mistakes made. Also while the dry run is taking place the existing hypotheses will be transformed from simple into complex hypotheses and decisions will be taken on the sort of variables for matrices and their order of entry when carrying out multiple regression analyses. International subject area committees will be involved in the finalisation of complex hypotheses and in decisions on their testing.

Thus it is hoped that in 1970 the full testing can be carried out with a minimum of errors.

It is hoped that the dry run for Stage 3 in 1970 will profit from the mistakes made for the Stage 2 dry run, and it may not therefore be necessary to hold a National Technical Officer meeting after the Stage 3 dry run, but this is as yet of course uncertain.

The bulk of the data processing requiring a large computer will be carried out at Columbia University. However, once the inter-correlation matrices are produced, it is possible that further work based on the inter-correlation matrices may be carried out at other centers, and in particular at Stockholm.

The assignments for writing particular sections of the report have not yet been made, but certainly the heads of the international subject area committees, together with selected members of those committees, will work in collaboration with the IEA Council members, members of the IEA central staff and IEA consultants in the writing up of the reports. The books will then be published commercially.

CHAPTER 11

THE DEVELOPMENT OF DATA PROCESSING PROCEDURES

11.0 Introduction

The first major problem has been the choice of an adequate computational facility for our work. The data processing requirements were well defined from quite an early date. We would need to handle files of data for three or four different age groups in each of about 20 countries. Each file would contain about 3,000 student vectors, 1,000 teacher vectors and 200 school vectors. A typical student vector would contain approximately 300 bytes of information, and the teacher and school vectors approximately 100 bytes each. The incoming data would be in a variety of formats, and its arrival at the data processing center would be spread over a period of between three months and 12 months. The analysis to be carried out after the organization of the date into sequentially indexed files would involve in general sequential processing with heavy amounts of computation (test scoring, correlation calculations, etc.).

It was provisionally decided to design the data processing system for the IBM System 360 computers, since these are available in most of the I.E.A. centers of activity, and indeed several member institutes of I.E.A. have their own System 360 capabilities. This range of machines is highly compatible, and in general the programs developed for use on one model would run efficiently with only a little modification on other models in the range. Particular 360 installations which were investigated in some detail included the 360/67 at the University of Newcastle in England, the 360/75 at Datacentralen in Stockholm, the 360/65 at University College in London, the 360/65 at the IBM Data Centre in London, and the 360/92 at the University of Columbia in New York. If possible, the analysis will be performed on this last-mentioned machine for the following reasons:

(a) It has superior peripheral storage devices and input/output facilities to any of the other systems investigated.

(b) It appears to be the cheapest ($300 to $400 an hour for CPU time).

(c) New York is conveniently located for I.E.A. as several of our part-time and consultant personnel are already working there.

The second major problem concerned the form in which data was to be transferred from the students who took the tests to the computer at the International Center. The possibility of having responses key-punched manually in each of the participating countries and the resulting Hollerith cards shipped to the International Center was rejected on the grounds of expense. Various optical scanning systems for use with marked answer sheets or cards were investigated. Of these, the best available appeared to be answer cards using the MRC 1501 scanner developed by Lindquist and his associates at the University of Iowa. One of these machines had been purchased by the Swedish government and was to be installed in Stockholm. Investigation showed that adequate time was available on this machine and that the cost per student would be two or three times cheaper than a system using the comparable IBM 1230. In addition, the answer cards had certain very obvious advantages over 1230 answer sheets. They were considerably less fragile and could be processed by any regular punched card equipment (i.e., sorters, collaters, etc.). It was also possible to pre-punch identification codes in the card, which would enable the student and particular test document to be uniquely identified, and these punched codes could be read by the optical scanning machine at the same time the scanning itself took place.

Eleven MRC answer cards were designed for use with Stage 2. Two of them are for use with Population I, seven are for various uses with Populations II, III and IV, one is intended for teachers and one is available for unspecified use if a participating institute needs a document on which to collect extra data. These cards are now being printed by IBM in Stockholm.

For certain data it was felt to be better to rely on manually key-punched cards for data transfer. This applies to certain questionnaire and test material for Population I (10 year old students, who may experience some difficulty in using answer cards) and to school information which is coded by the school principal in a questionnaire booklet. National Centers who doubted the practicability of using answer cards in their country were given the alternative of sending in all their data on key-punched cards, and at least one country will be exercising the option during our dry run activities. As a result of our experiences in the Dry Run, this National Center, and perhaps others, may wish to change their procedures for the main testing in 1970.

Below we give in as much detail as is currently available, the planned sequence of data processing operations for handling data of one particular country. The numbers given in the right-hand column refer to the time in months after the actual administration of the tests in a particular country. During 1969, the month of administration is likely to range from February to October, and in consequence the dates at which other stages of the data processing are reached will have a nine-month spread.

11.1 Timetable

	Months After Date of Testing
Step 1 National Centers send to the central staff precise details of their sample. This consists of: (a) the code number for each school (b) the number of students to be tested in each school (c) the number of teachers to be given questionnaires in each school (d) a list of the tests it is proposed to administer	-2
Step 2 Central staff prepare punched MRC cards to correspond to these specifications. This will probably be done on the N.F.E.R.'s 1130 computer at Slough in England. The identification codes will be interpreted onto the top of each card, and then the cards will be sorted into the appropriate order for use by the National Center (all cards for one student and then all students for one school will appear together in the pack). The cards are then sent by the fastest available means to the National Center.	-1
Step 3 National Centers carry out the administration of the testing in the schools which have fallen within the sample.	0
Step 4 National Centers send completed MRC cards to Stockholm for scoring.	+1

	Months After Date of Testing
Step 5	
National Centers send punched card data direct to the processing center in New York, retaining a copy of the data for themselves to guard against accidental loss or damage in transit.	
Step 6	
The MRC cards are read on to tape by the 1501 machine in Stockholm. Each card type will go on a separate tape. In addition to some checking of the data, the items will be re-ordered at this stage into the sequence that we require rather than that in which the data from the card is actually read. Duplicate tapes are made, one for each card type, and a copy sent to New York.	+2

Step 7

The main data processing begins with the reception of the punched cards and magnetic tapes in New York. It is proposed that at first each of these be checked on the small Teachers College computer to find out what each data set contains and whether or not there have been any major errors. If errors are disclosed at this stage, the data will be referred back to the appropriate people in Stockholm, Hamburg or the individual National Center.

Step 8

The incoming data sets, now all on magnetic tape, will be processed on the main computer at Columbia (360/92). Five files will be created:

MF: master file containing all international student data

RT: raw teacher data file - this will contain all school questionnaire data, together with the MRC card type 3 data which contains the school ratings of opportunity to learn the items in the Science test

Month After
Date of Testing

RS: raw school data file - this will contain all school questionnaire data, together with the MRC card type 3 data which contains the school ratings of opportunity to learn the items in the Science test

SS: special Science file containing response data for the IVS Science specialist tests which are being given in some countries. It is intended that this data will be subjected to some international analysis to answer questions principally concerned with curricula

NO: national option file containing all other data collected. It is not intended that this file will be included in the international analysis, although data will be available for national analyses after the main international analysis is completed.

There will be some 60 master files, one for each country and each population. Once they are complete, they will be stored on magnetic tape (maximum size of file estimated at 70,000,000 bytes), but during the creation of the files it is anticipated that disk storage (2311 or 2314) will be used extensively. The files will be organized as follows. Each student will have one record which will be the sum of up to 8 input records (MRC cards) or 13 input records (punched cards). All the students in one school will be stored together in ascending sequence of student identification number. Schools will be stored in ascending sequence of school code number. There will be 3 possible record lengths for each student within any one population, depending upon whether a country has administered (a) General + Science tests, (b) General + Science + Reading Comprehension tests, or (c) General + Science + Reading Comprehension + Literature tests. There are slight variations in record length between populations because the test instruments have slightly different numbers of items. +4

Step 9

While the files are being completed, extensive editing and checking will be carried out to ensure that the

Months After
Date of Testing

variables all lie within the correct ranges and that the data is essentially complete. Once the files are complete, a census will be taken to find out exactly how many records there are of each type for each stratum in the sample. Reports of this census will be sent to National Centers. +5

Step 10

Data containing details of the population and its division between the various strata will be combined with the results of the census to calculate sample weights for each student and each school.

Step 11

The tests will be scored and certain other variables will be derived from combinations of raw variables in the master file.

Step 12

Two working files (WF1 and WF2) will be produced from each master file, i.e., one for each population-country combination:

WF1 contains school variables, school means of teacher variables, and school means of student variables. This will contain about 120,000 bytes for one file (or about 8,000,000 all told). This could be stored with ease on one 2314 disk pack.

WF2 contains student variables drawn from MF. Individual item responses will not be retained in this file, and are replaced by test scores, sub-scores and other derived variables (400,000 bytes for one file or about 24,000,000 all told). This can be stored on one 2314 disk pack.

Step 13

Item analysis of all cognitive and attitude scale data, including principal components analysis.

3

Months After
Date of Testing

Step 14

Computation of marker variables from sample (by stratum), then the calculation of univariate statistics (means, standard deviations and frequency distributions) for all variables using WF1 and WF2. These results will be reported in full to National Centers. +6

Step 15

Initial screening of variables to decide which of them will be included in the main correlation matrices. The screening will be conducted in the following terms:

(a) the size of the simple correlations with the main criterion variables.

(b) the interest a variable or pair of variables may have for testing of specific hypotheses.

Step 16

Calculation of bivariate statistics (simple correlations) for two sets of data, (a) "between" schools (WF1) and (b) "within" schools (WF2 - WF1).

Step 17

Screening of variables for regression analyses. Variables will be chosen from the correlation matrix for inclusion in regression analyses in terms of their intrinsic interest and the size of their simple correlations with one or more of the criteria. The variables so selected will be grouped together according to the type of variables (e.g., home background).

Step 18

Regression analysis. This will be carried out to predict the various criterion scores from the other variables. The variables will enter the regression equations one group at a time and estimates of group contribution to the explained variance will be made in addition to calculating estimates of the individual contributions of variables.

Months After
Date of Testing

Step 19

Computer analysis required to test hypotheses which were not covered by the regression analysis.

Step 20

Regression analysis and other hypothesis test results are reported to National Centers for their consideration. Any necessary re-runs of these tests are carried out. +7

Step 21

Where necessary national data files will be reconstituted to include national option data. The files will be put into a format acceptable to the National Center and the data will be returned for national processing.

Step 22

Conclusion of main data processing for one country. +9

The concluding stage will be reached between December 1969 and August 1970 for the Dry Run, depending upon the country with which we are concerned. As soon as Step 12 has been concluded for a sufficient number of countries, a between countries principal components analysis will be conducted on the test data to verify the weighting of the different sub-scores that go to make up the total score. This will certainly be needed for Science, and may also be needed for Reading Comprehension and Literature. If necessary, sub-score weighting will be revised for the main testing.

The results of the item analyses, the univariate statistics, and the comments gathered from National Centers during the Dry Run exercise will be used to undertake a revision of the entire test instrument battery by October 13, 1969. A meeting of the National Technical Officers will be held in January 1970 to settle any remaining problems concerning procedures at National and International levels. Special discussions will be held with those National Centers who are planning to test very early in 1970 for whom the discussion at the National Technical Officer meeting would be too late.

The work for the next 12 months, i.e., for the full testing, should then follow in the exact sequence outlined above.

CHAPTER 12

12.0 POSSIBLE CONSEQUENCES OF IEA STUDIES ON EDUCATIONAL RESEARCH AND EDUCATIONAL POLICY DECISIONS

The IEA approach to cross national research has been in existence for approximately ten years. The essence of this approach is a cooperative attack on a series of educational research problems by a collection of National Research Centers utilizing a central coordination staff, centralized data processing, and expert committees and consultants drawn from the participating countries. Although the IEA studies to date have been based on the use of tests and questionnaires and quantitative methods, it is clear that IEA research procedures need not be limited to such methods where the problems being investigated can be better studied by other methods.

Because of the complexity of the problems it has attacked so far, the IEA group has been relatively cautious at each step of the research. As a result, three to five years have been required for each complete study. Hopefully, it will be possible to reduce the time for studies in the future while still maintaining the research rigor and cooperation of the participating National Research Centers.

The pilot study of the IEA (Foshay, 1962) was a study of the feasibility of doing cross national research with international tests and questionnaires. This study demonstrated that international instruments could be developed and that, with appropriate care in translation and provision for recognizing cultural and national differences on specific questions, such instruments could yield comparable and meaningful results. The pilot study revealed the great interest of the participating National Research Centers in this type of research but it also made clear many of the problems of finance, sampling, administration, and interpretation which had to be solved at the national level. In a number of cases, particular National Research Centers were found inadequately staffed to participate in a cooperative cross national type of educational research.

The problems posed in the pilot study were more carefully attacked in the Mathematics project. A number of centers could not participate because of inadequacies in staffing and finance and they had to be replaced by stronger National Research Centers. Flaws in sampling during the pilot study had to be corrected by the use of more rigorous sampling procedures under the direction of an international sampling expert. Detailed procedures for the development of specifications for internationally valid tests required far more time and care than had been originally anticipated. Empirical tryouts of test material and questionnaires revealed many flaws in the materials which required modification.

One of the consequences of the IEA procedures so far is a clearer understanding of the problems of constructing internationally valid tests and questionnaires. Procedures involving specialist groups in each country, development of national papers, development of international specifications followed by careful and imaginative test construction, criticism at the national level, and empirical validation of test material are required for the tests to satisfy both empirical as well as content validity for the many national groups involved. We believe that in spite of the extraordinary costs in time and money, the development of internationally valid research instruments has enormous consequences for educational research and educational planning.

In the Mathematics study (and hopefully in the present study) these instruments reveal the levels of achievement which have been attained by students in various school systems. The diversity between national systems as well as within these systems has made it possible to secure a fresh look at each system against the background of all the other systems. From the national reports so far published it is evident that each country is concerned about its present output of mathematical competence and each is seeking ways of improving education in this area. Most of the countries plan innovative procedures to correct what they believe to be inadequacies and look forward to the use of parallel mathematics tests at a later stage to determine the effect of the educational changes instituted so far.

Some of the nations have been concerned about dual systems of education which attempt to develop a small educational elite in contrast with lowered expectations and educational opportunities for the large majority of students. The IEA reports, both international and national, have highlighted the differential effects of dual systems and comprehensive systems of schools. Clearly, further evidence will be necessary before massive changes in the educational systems should be instituted. It is to be hoped that the current IEA studies will shed further light on these problems and serve to help educational policy makers in the different countries make more realistic decisions on the basis of evidence as well as national values.

Since each nation supported its own participation in the IEA, each nation regards its portion of the study as being of vital importance to its decisions about educational policy. It is this cooperative attack on common educational research problems by approximately twenty National Research Centers which represents a new dimension in educational research. This is in sharp contrast to a highly centralized study where the results are regarded as those developed and interpreted by some small group of international research experts. Thus, it is believed that the IEA finding for a particular country would be reacted to in a defensive way if the nation thought of the researchers as foreign and without a thorough understanding of its own situation. We have found that each country accepts its own IEA results as being relatively valid --

especially where its own major research group is regarded as competent and objective. The national reports, written by the nation's own researchers, have clearly been accepted as relevant by each of the national groups and efforts are under way to correct some of the educational problems pointed up in these national reports. What is emerging is the original thesis that each nation may learn from the experiences and results of the other nations -- where it regards the evidence as relevant to its own national system of education. Thus, we have attempted, with some modest success, to move to a position where the entire world may be regarded as a laboratory for educational research. What is true of other nations may be one basis for the modification and improvement of each national system of education.

The IEA has already, and is likely to yield further evidence on the characteristics of schools, teachers, curriculum, and instruction which have consequences for both cognitive and affective outcomes in selected school subjects. The relations between what is taught and what is learned have emerged very clearly in the Mathematics study. It is to be hoped that this will be equally clear in the school subjects under consideration in the present study. These studies reveal most clearly where the schools may be modified to secure better results (if this is desired). Although additional evidence may be needed on this point, the studies also suggest the characteristics of schools, teachers, and of instruction where modification is least likely to produce significant differences in student learning.

The IEA Mathematics study was not as successful as expected in revealing the home and community characteristics which influence school learning. However, the research work on the home and community done by many researchers in the period from 1964 to 1968 have suggested more effective techniques which are being implemented in the present study. It is to be hoped that these techniques will be more effective in separating school and instructional variables from the effects of home and community variables in the present investigation.

In the research on home, community, school, and teaching variables we are attempting to distinguish between malleable and less malleable variables. If this attempt is successful, it is clear that educators and educational policy makers should concentrate their efforts on altering the malleable variables where the evidence is clear that they have a significant effect on the cognitive and affective outcomes of student learning.

The IEA findings have important implications for who shall be educated. It is possible to analyze the results to determine the consequences of comprehensive versus dual systems of education, the consequences of additional years of education for various groups of students, and the consequences of different arrangements of schools, student grouping, and selective processes on the educational yield of particular countries. These are basic questions which each nation must answer for itself. The IEA can only provide some of the data showing the consequences of various decisions in selected countries.

The IEA has attempted to involve economists, sociologists, political scientists, psychologists, and comparative educators in planning the data gathering as well as the data analysis. The economists, especially, have suggested improved methods of studying cost-benefit data for education. It is anticipated that our present study will better enable us to understand the ways in which national educational characteristics can be investigated in relation to learning outcomes. While these are difficult areas to investigate, we believe that our present efforts will make a modest improvement over previous methods of relating the educational differences among the nations to the learning outcomes evaluated by our tests and questionnaires.

Our inquiry is limited to selected subjects and to selected countries. However, our larger goal has to do with an improved understanding of the factors which account for educational outcomes at the student and school level. This is a long-term inquiry and we hope that each IEA study will yield some gains in our progress toward a fuller understanding of education and how to improve it.

REFERENCES

"Cross National Conference on Education, Manpower, and the Economy" New York: Edited by D. Super. Teachers College, Columbia University. (In preparation)

De Landsheere, G., "Une recherche qui fera date: Mesure du rendement de l'enseignement des mathematiques dans douze pays." Belgium: La Nouvelle Revue Pedagogique, Tome XII, No. 10, June 1967.

Foshay, A.W. (Ed.), *Educational Achievement of Thirteen-Year Olds in Twelve Countries*. Hamburg: Unesco Institute for Education, 1962.

Hotyat, F., *Les Examenes: les moyens d'evaluation*. Paris: Bourrelier, 1962.

Hultin, Mats. *Skoland Produktivität.* Sweden: Almqvist & Wiksell, 1968.

Husén, T. (Ed.), *International study of educational achievement in mathematics: A comparison of twelve countries*. Volumes I and II, New York: John Wiley & Sons, 1967

Keeves, J.P., *Variation in Mathematics Education in Australia*. A.C.E.R., 1968.

Kumbaraci, T., Unpublished Ph.D. thesis, Teachers College, Columbia University, 1966.

Kuusinen, Jorma, *School Systems and Level of Achievement: Some Notions on the Basis of the IEA Study in Mathematics*. Finland: University of Finland, School of Education, February, 1967.

Kuusinen, Jorma, *International Study of Achievement in Mathematics*, Finland: Report from the Center for Educational Research, 1967.

Lyttkainen & Pavliinen, *Finland: A National Analysis of the IEA data*, International Review of Education, Vol. XV, No. 2, 1969.

National Institute for Educational Research, *National Report of the IEA Study in Mathematics*, NIER, September, 1967.

Pidgeon, D.A., Achievement in Mathematics: *A National Study in Secondary Schools*, the Mere, Upton Park: N.F.E.R., 1967.

REFERENCES (cont'd)

Stern, H.H., (Ed.) Foreign Languages in primary education: the teaching of foreign or second languages to younger children. Report on an international meeting of experts, April 9 - 14, 1962. Hamburg: UNESCO Institute for Education, 1963.

Stern, H.H., (Ed.) Language teaching and the Young School Child. Oxford: Oxford University Press, 1968 (in press).

Wall, W.D., (et. al.), Failure in School, Hamburg: Unesco Institute for Education, 1962.

Wolf, Richard M., Data Bank Manual, International Project for the Evaluation of Educational Achievement. Sweden: Almqvist & Wiksells, 1967.

APPENDIX A

PROPOSALS TO IEA FOR FUTURE WORK

Memorandum

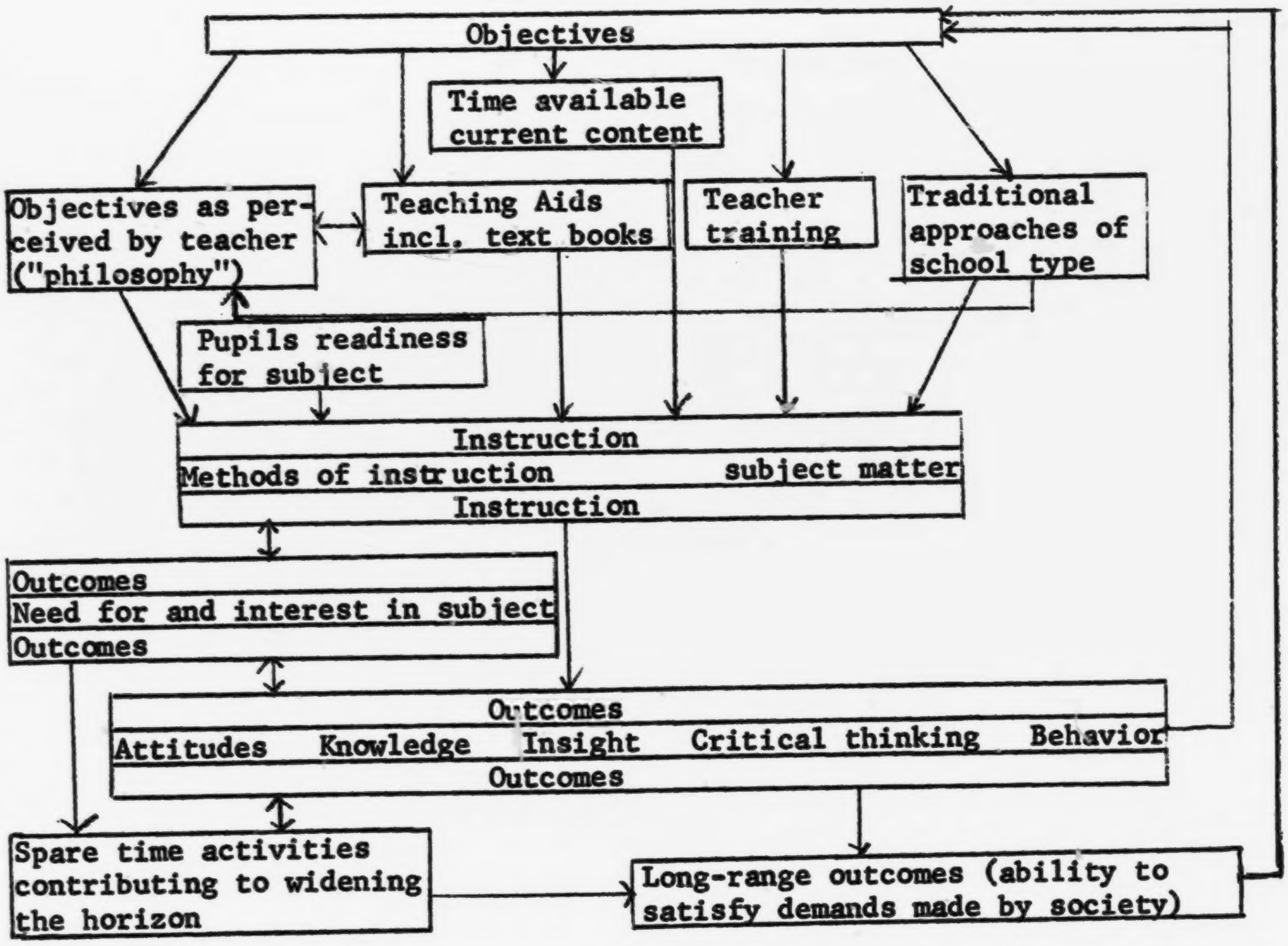
Proposal for an International Investigation of Instruction in Civics

Prepared by
B. Bromsjo and Committee
Stockholm

Starting-point

When education authorities introduce a new subject, they do so with definite ends in view. The intention is to achieve results in relation to the objectives set up for the subject. A certain number of teaching periods is allotted to the new subject, but the results achieved depend upon many so-called intervening variables, which must be taken into consideration when the results of instruction are to be judged. The model shows some important factors connected with the subject civics.

FIGURE 1


Objectives
Time available current content
Objectives as perceived by teacher ("philosophy")
Teaching Aids incl. text books
Teacher training
Traditional approaches of school type
Pupils readiness for subject
Instruction
Methods of instruction
subject matter
Instruction
Outcomes
Need for and interest in subject
Outcomes
Outcomes
Attitudes
Knowledge
Insight
Critical thinking
Behavior
Outcomes
Spare time activities contributing to widening the horizon
Long-range outcomes (ability to satisfy demands made by society)

The intervening variable, which may be most important for the success of the subject, is instruction. This may be divided into methods of instruction and subject matter. But this variable, about which the authorities give advice and instruction, is in its turn affected by teaching aids, including text books, the teacher (training and "philosophy") and the tradition which may be formed in certain cases. (See Figure 1)

Some spare-time activities may also be regarded as results of the work of the school. These may, in their turn, influence the result of instruction in several ways.

The school endeavours in the last analysis to achieve long-range results, called in the model ability to satisfy the demands set up by society.

Proposal for an investigation

It would naturally be of prime interest if the projected IEA investigation made ability to satisfy demands set up by society its central task. But in this we are faced with great problems, arising out of lack of evaluation instruments. It will also be difficult to keep the variable called spare-time activities under control, since such activities are not purely the result of school work. Hence it seems necessary to limit the study to the variables which can be assigned more directly to the instruction provided by school.

The result variables suggested as subjects of evaluation are:

A. Attitudes Attitude to minority groups, civil phenomena, the school, religion etc.
B. Interests Different school subjects, various aspects of civics, different forms of work.
C. Critical thinking Evaluation of various kinds of information. Reasons for holding certain opinions etc.
D. Insight Causal connection. Importance of geographical and biological conditions. Historical aspects (e.g. economic problems, different kinds of intervention by society, development of society).
E. Knowledge The course studied.

The parts of the subject which may be of interest in this connection are:

1. Problems of co-existence (The function of the small group.)
2. Purpose of society. Rights and duties.
3. Working life. Occupations and choice of occupation.
4. Social problems. Social policy.
5. Law and justice.
6. Economic problems.
7. Democracy and problems of democracy (including propaganda, political parties, organizations, formation of opinion).
8. International problems.

In accordance with this, the investigation model would be as below in respect of the result variables:

	A	B	C	D	E
	Attitudes	Interests	Critical Thinking	Insight	Knowledge
1. Problems of co-existence					
2. Purpose of society					
3. Working life					
4. Social problems					
5. Law and justice					
6. Economic problems					
7. Problems of democracy					
8. International problems					

The investigation would refer to pupils aged about
I 13 years
II 16 years
III 19 years

Investigation instruments: Tests. Interests and attitude inventories. Possible types of items are given in the following pages.

The general plan of the investigation should allow of the special treatment of differences in attitudes, interests, knowledge etc. between the sexes.

It should be further observed, possibly in connection with some other subject being investigated, to what extent the instruction given at school brings the pupils into contact with material in which other countries and people are described as curiosities, that is to say, strange customs and traditions are stressed without any background or explanation being given.

Since it is then of interest, naturally, to give the background of the results of instruction, it seems urgent to collect data referring to this background. It is therefore suggested that the teachers of classes affected by the investigation shall answer a questionnaire.

The teachers may be asked about

1. The objectives of instruction.
2. Methods of work applied.
3. Tuition time devoted to the various items.
4. Conditions of work (material needed, views on material for different kinds of pupils, textbooks etc.).
5. Their own training and other background factors.

Some of the types of items which could be included in the international investigation in civics are suggested in the following material.

Pupils interest in different subjects

Below are a number of school subjects, and after them are squares in columns. Above the columns are the headings "Very interesting", "Rather interesting" and so on. Make crosses in the squares to show what you think of the subjects. If you think a subject is neither interesting nor dull, put a cross in a square under the heading "Neither interesting nor dull". Show what you think about all the subjects you have at school.

	Very interesting	Rather interesting	Neither interesting nor dull	Rather dull	Very dull
Mathematics	☐	☐	☐	☐	☐
Civics	☐	☐	☐	☐	☐
"	☐	☐	☐	☐	☐
"	☐	☐	☐	☐	☐
"	☐	☐	☐	☐	☐

etc.

Pupils attitude to different methods of work
(Harnqvist's model)

Which do you prefer? And which is most usual now?
Below are different descriptions of work in school, arranged in pairs. For each pair you must make a mark (a cross) to show which you yourself prefer. You must also show which method is, according to your own experience, most usual in school work now.

I prefer this	This is most usual now	
....	....	That the teacher deals with subject matter in the same order as it appears in the textbook
....	....	That the teacher deals with subject matter in different order than that in the textbook

....	....	Short questions on facts
....	....	To give an account of a larger section of the homework

A 3: I, II*

Eve is very clever at school. Her best subjects are mathematics, physics and drawing. She is good in all other subjects, too, except needlework, which she does not like, and gymnastics. She is somewhat delicate, and her legs are weak. She is diligent, careful and independent in her work, but rather shy and reserved. She is very keen on photography in her spare time, does her own developing and printing, and mounts the prints herself. Which of the following occupations do you think she ought to choose?

---

*The code refers to the behavior (A, B, C, etc.), the content (I, 2---8) and the population of students (I, II, etc) as shown in the table on page A-13.

○ Clerk
○ Mechanic
○ Engineer
○ Child nurse
○ Chauffeur

A similar item for male pupils.

A 4: I, II, III

With which of the following statements do you agree? (Only one may be chosen.)

○ The state should pay all the costs when a person is ill
○ The state should pay most of the costs when a person is ill
○ State aid should be dependent upon the income of the person who is ill
○ The state should pay all the costs for people without money
○ The state should not interfere in problems of health

A 5: II

A sixteen-year-old has been guilty of five thefts of motor-cars, three burglaries in shops, and one housebreaking into an empty weekend cottage. He has never been charged before. What do you think his sentence should be?

○ Remission of prosecution
○ Two months' probation
○ One month's imprisonment
○ Three months' imprisonment
○ Twelve months' imprisonment

A 7:I

Which of these views is most like yours?

○ Only girls should help with housework
○ Girls should help with housework more than boys
○ Boys and girls should help with housework as much as each other
○ Boys should help with housework more than girls
○ Only boys should help with housework

A 7:I

Which of these views is most like yours?

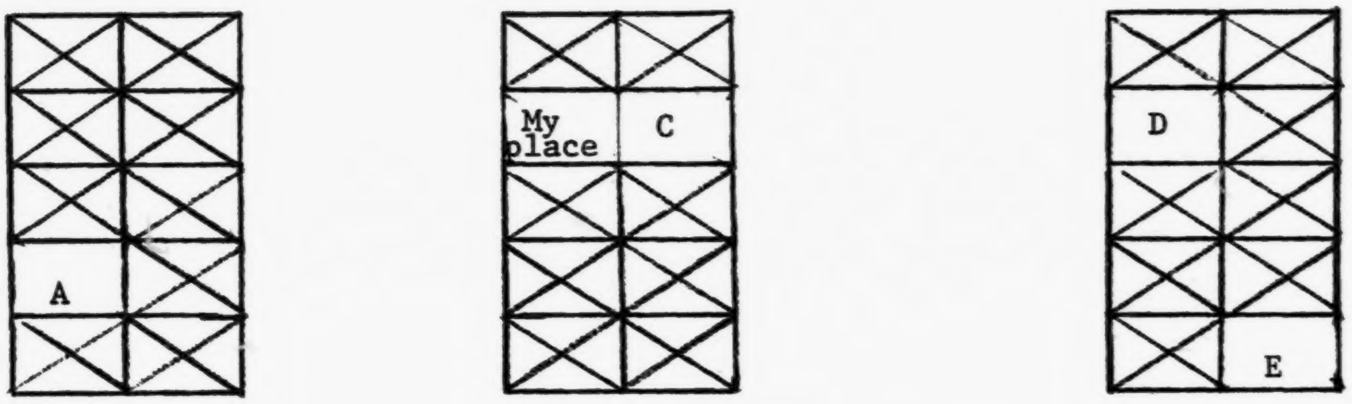
○ Pupils of your age should help with all housework
○ Pupils of your age should have certain fixed tasks at home; e.g. make beds, run errands and the like
○ Pupils of your age should help with housework when they have finished their homework
○ Pupils of your age should help with housework when they have time
○ Pupils of your age need not help with housework

A 8: I, II, III

There is a newcomer in your class, who is (is from) .... . Where in the class do you think this new pupil should sit?

☒ = occupied place   Notice your own place.

Teacher



O A
O B
O C
O D
O E

C 1:III
Catchwords and slogans are often used--in politics, for example. They are hardly ever wholly correct or true. One of the statements below is quite false. Which?

O Politics is the art of doing what is possible
O The world is a powder-magazine which may explode at any moment
O Propaganda is necessary
O As long as people negotiate they do not fight
O The World War brought peace to the world

C 2:III
If the highway system is the responsibility of the state, it does not follow that the state should

O Issue general traffic regulations
O Acquire money for maintenance
O Manufacture vehicles
O Punish those who violate traffic regulations
O Close highways when necessary

C 5: II, III
Juvenile delinquency has increased greatly since the end of the war, particularly in towns. What conclusion can be drawn from this?

O Peace led to increased crime
O Young people in towns are "worse" than young people in the country
O The police are more efficient in the towns than in the country
O Young people are "worse" now than formerly
O None of the above conclusions can be drawn

C 6: III
Since the end of World War II (1945), prices of many goods have been more than doubled. Mark with a cross the statement below which supports this claim.

○ All people earn more than in 1945
○ It is getting more difficult to make ends meet
○ People do not buy as many goods now as they did twenty years ago
○ Money deposited in banks has sunk in value
○ The quality of goods is more than twice as high as it was twenty years ago

C 7: II, III

One of the following statements is not included in the UN declaration of human rights. Which?

○ Everyone has the right to life, liberty and security of person
○ Everyone has the right, in the interests of the UN, to ignore a country's laws
○ No one shall be subjected to arbitrary arrest, detention or exile
○ No one may be compelled to belong to an association
○ Everyone, without discrimination, has the right to equal pay for equal work

C 8: I

One of the following statements is false. Which?

○ Women are suitable for engineering work
○ There are no illiterates in the world now
○ Negroes are as gifted as whites
○ India has had Nobel Prize winners
○ In countries with a high standard of living, children go to school for a relatively long time

C 8: II

Choose the best answer to the following statement.

Statement: We know that our ambassador to X-land will return home if diplomatic relations are broken off with X-land. If we know that our ambassador to X-land is on his way home we can

○ Be sure that diplomatic relations with X-land have been broken off
○ Be sure that diplomatic relations with X-land have not been broken off
○ Not say, from the information we have, whether diplomatic relations have been broken off or not
○ Assume that it is very likely that diplomatic relations with X-land have not been broken off
○ Assume that it is very likely that diplomatic relations with X-land have been broken off

D 3: 1

Which of the following pieces of information ought one to obtain first about the occupation one is thinking of taking up?

○ Length of training period
○ Demands made by the occupation
○ Level of income
○ Hours of work
○ Holidays

D 4: I, II
In most countries, the situation of women in the labour market is less favourable than that of men. Why?

O Old prejudices remain in the labour market
O Women's performance are inferior to men's
O Women are more interested in housework
O Women are not strong enough
O Men have to support families

D 6: II, III
Risks of higher prices become greater if people

O Save more
O Buy more
O Import more
O Export more
O Produce more

D 5: I, II
What do you think is the main reason for punishing criminals?

O To protect the public from the criminal
O To make the criminal feel ashamed of himself
O To deter other people from committing crime
O To help the criminal to become a good member of society again
O To make the criminal pay his debt to others

D 7: II
Below are five statements made by different people. One of them may be said to express a democratic attitude. Which?

O Children should obey their parents without question
O All people cannot share in the government of a country
O Decisions should be made by ballot after everyone has been allowed to express his views
O The state, not the citizen, decides what is best for the country
O The task of employees is to do as they are told, and nothing more

E 2: II
What is parliamentarianism?

O The principle that the government is chosen by the party with a majority or the greatest influence in parliament
O A democratic system of government with, among other things, universal suffrage at parliamentary elections
O That meetings of parliament are public
O The principle that important decisions in the government and parliament are preceded by negotiations between the largest parties

E 6: II, III
A newspaper wrote: "In most western European industrial countries there has been gradual inflation ever since the end of the war". What is meant by inflation?

O Drops in wages
O A fall in the value of money
O Purchasing power increases
O Prices of goods drop
O The supplies of raw material decrease

E 8: II, III
Which of the following UN organization is concerned with problems of agriculture?

O WHO
O UNESCO
O ILO
O FAO
O UNICEF

Testing the Ability in English as a Foreign Language

A report by the Finnish subcommittee.

I Principles of Classification

Language represents a system of communication and the language skills can be defined as habits of communication. The messages are usually transmitted along auditory and/or visual channels. The transmission can take place by means of lingual or extralingual signs or cues. Phonemes are the basic elements of the communication in the auditory channel and graphemes on the visual channel. (In language testing only the graphemes represented by letters are usually included.) In addition, extralingual material is made use of: the communication comprises other visual cues (gestures, expression, acts) as well as other auditory cues.

1. Perceptual-motor and meaning components of language

In the examination of language abilities a distinction should be made between perceptual and symbolic processes. Perceptual or perceptual-motor functions represent the lowest level of integration. The meanings are, however, essential for a language. The meanings contain a symbol relation: they represent some extrapresent object called significate (the level f mediating processes). The concept of meaning can be defined as a relationship between a significate and its sign. Regardless of the way in which the relationships between the significate and the sign are interpreted in detail, the criteria for the recognition of the symbol relation should be settled, if the process of learning a foreign language is studied since the learning processes for perceptual-motor reactions differ from those for the meaningful associations. Concept formation is a necessary condition for a mediating process. Pure perceptual associations occur, e.g. between written and spoke languages, when there is no recognition of symbol relation.

The sounds and their arrangements into larger units with definite forms and distributions may be learned without symbol relations by young children and, to some extent, by adults, at the early stage of learning. Usually, however, these elements of the structure of language are learned in close connection with meanings. In language testing only the "public" meanings (especially the linguistic meanings) which are common, constant elements of communication to most members of the large community are assessed through language testing.

2. Uncoding and decoding

Both the identification and discrimination of percepts and the comprehension of meaningful symbols are acts of recognition or "decoding", while the imitation and reproduction of perceptual elements and the delivering of meaningful symbols can be regarded as acts of production (encoding). These functions are distinguished in a test situation: in connection with production the persontaking the test always reacts by using symbol language (by speaking or writing), whereas recognition which consists of listening or reading processes can be indicated by simple extralingual choice reactions.

Classification scheme for language abilities

Level of integration	Channel	Recognition	Production
Level 1	Auditory	Identification and discrimination of sounds	Pronunciation, intonation and stress
	Visual	Identification and discrimination of letters	Spelling, punctuation and capitalization
Level 2	Auditory	Recognition of words Word and sentence patterns	Active vocabulary Production of grammatical units
	Visual		
Level 3	Auditory	Auditory comprehension	Speaking
	Visual	Reading comprehension	Writing

In the classification scheme, the hierarchy of the levels of integration is determined by the complexity and the number of symbolic processes required. In the study of foreign language some categories may be of minor importance, e.g. the identification and recognition of letters, provided that the subjects are able to read.

Some psychological processes cannot be described within this scheme. The direction of the associations between the native and the foreign language (foreign->foreign, native->foreign, foreign->native) is important for the study of learning in some categories. The interactions between the auditory and visual channel should be considered in connection with certain processes.

In addition to these psychological processes, the structural elements of language which constitute the grammar, should be classified. The grammatical material of languages can be divided into units: phonemes (graphemes), morphemes, words, compounds, phrases, clauses, sentences, and sentence sequences constitute these classes at different hierarchial levels. These divisions can be made in different ways in different languages. The grammatical categories of English language are the most natural starting points, when a taxonomy is to be compiled for countries where English is not the native language. At a low level of the hierarchy, those habits of articulation, syllable types, and collocation are included which are related to patterns of intonation, stress, and rhythm. At higher levels, the patterns of words (words, modifications, sequences) are considered.

4. Associations among percepts and symbols

Language ability is psychologically based on association chains. The associations can be divided into two main groups. The first comprises associations concerning the overlap between words or concepts within one channel of communication. The second comprises the corresponding associations among different channels of communication. The associations based on the structural units of language can be separated from both of these. The similarity relationship only is discussed in detail.

Similarity relationships. Associations indicating the degree of similarity may refer to both the perceptual and the meaning level. A similarity of meanings implies an equivalence relationship. A partial similarity may refer to a subsumption or a partial overlap relationship. Rather common in language tests are, e.g., the following synonymity assertions: "Word A has the same meaning as word B." "Word A has the same meaning as its definition B." Relationships indicating the degree of similarity can be substituted by a dimension ranging from similar to opposite meanings (synonymity - antonymity, + - -) or by a (half-) dimension from similarity to neutrality (+ - 0). If we use the models originally introduced by Osgood, in which the word or sentence meanings are measured by adjective scales, we may expand them to a more general scale, e.g. substantive antonymities, antonyme sentences etc. at the extremes of the scale (+ --) or any words or sentences at the extremes of the scale(+ - 0).

Many other techniques can be used for examining the associations which indicate the degree of similarity. In differential psychology it is customary to study similarity relationships by means of synonymity tests, by word grouping or word analogy tests. The basic sentences showing similarity relationships may be transformed into interrogative sentences etc., but they always are included in the instruction of language tests irrespective of which techniques are applied for the assessment of perceptual or meaning associations.

5. Problems of learning a foreign language

The learning tasks presented to the students differ according to the nature of their native language. Therefore, cross-language comparisons provide in this case a unique problem for research, going beyond the problems concerned with most school subjects. The process of learning is closely connected with the amount and the kind of similarities between the pairs of language (native/foreign) in question. It should be possible to apply general principles concerning the effect of similarity of material (stimuli and responses) to cross-language comparisons.

The amount of positive and negative transfer of the habit systems can be predicted by systematic linguistic comparisons of the pairs of languages. It may not, however, be possible to predict in advance, what elements of the language structures and contents are essential for the learning of global habits of communication in a new language. It should be examined, e.g., to what extent the phonetic similarity (variants of one phoneme or different phonemes) is related to the distinction between phonemes or to the production of phonemes. The same question can be presented with regard to the other elements of the language structure and content, e.g, the similarity of intonation, the similarity of word sequences or sentence patterns, the amount of overlap in vocabulary etc.

A linguistic examination of similarities between languages is, however, not sufficient for the purpose of cross-language comparisons. Therefore, we should examine whether different languages differ from each other in the frequency of the occurence of some structures and contents in such cases, when they have similar characteristics. In a more general way, an examination of the frequency of presentation of various types of learning tasks would be important for the comparison of achievements in languages.

Since the directions of the association chains (native/foreign) seem to be very important factors in language ability, the process of learning should be examined in detail from this point of view.

6. The categories for testing ability in English (as a foreign language)

The choice of categories for language tests must depend on the nature of the problem studied. For a global assessment of achievement it might be sufficient to include a language test with subtests for four skills: speaking, listening, writing, and reading. For the problems of learning presented above, as well as for more detailed diagnostic purposes, it would be necessary to include both global and more restricted ("analytical") variables in the battery of language tests.

The following remarks refer to the classification procedure.

1. Language is regarded as a system of communication. Therefore, the definition of grammatical concepts etc. has been omitted, although these concepts have been applied to concrete tasks.

2. Some content areas have been analysed carefully, but these classifications have not been included in the report. They would be relevant to the comparisons of curricula as well as in test construction.

3. The principles of classification vary to some extent in various main categories, because the psychological processes and the technical requirements of testing have been considered simultaneously. Ex. the "partial production" techniques have been discussed separately.

4. The system of categories has not been put in the form of Bloom's general taxonomy, although many parts of it can be easily presented in Bloom's terms.

II Classification

1. Pronunciation

Identification and production of vowels and consonants in different contexts.

Recognition of sound segments and discrimination between them without recourse to any specific meaning.

The sounds to be recognized are usually presented as parts of words and sentences from the tape.

Ex.: Similar or different? (1) sleep - slip (2) fist - fist? (3) jeep - gyp

Recognition of sound segments and discrimination between them through comprehension of meaning.

Ex.: The teacher reads a word or a sentence and the students choose from two or more pictures the one that fits the words or sentences.

Production of sound segments in words

Ex.: The production of correct sounds can be recorded in a situation where the student has to answer questions or to react to other kinds of stimuli by using words which contain problems of pronunciation.

Indirect recording of the production of sound segments.

This technique is also called "partial production". The student is asked if a word that is given to him in writing is the same as the sound of another word which is also presented to him in writing.

Ex.: Compare the sounds represented by the vowels. Write the numbers of the sounds that are the same.

sea (1) great (2) head (3) cheap (4) breath

2. Stress (Accentuation)

To give prominence to some syllables and words over others. Difference in prominence may depend on loudness, duration, and pitch.

Recognition of the syllable receiving a particular stress in words.

Stress cannot be totally distinguished from intonation in these tasks.

Ex.: The teacher reads aloud a word and the student is asked to mark the most prominent syllable(s).

Production of words with problems of stress.

The general technique is to ellicit from the student the words or sentence containing stress problems. Scoring would be by direct observation or by recording.

Indirect recording of the production.

The techniques consist of written words or sentences which contain stress problems. The student is asked to mark the syllable with the primary stress etc.

Ex.: (1) af-ter (2) with-out (3) al-ways (4) be-fore (5) be-tween

3. Intonation

Sequences of pitch in the speech. Normal statements; falling and rising tune.

Recognition of the intonation without recourse to any specific meaning.

Ex.: Compare the intonation, not the words, of the three sentences you will hear. Write the numbers of sentences that have the same intonation.

Identification of the intonation through direct meaning of the intonation.

There are only some meanings which are closely and consistently related to the intonation.

Ex.: Listen to the intonation. Check the choice that best fits the meaning. - Organize an export business? The speaker is probably (1) making a report (2) making a request (3) asking for confirmation.

Production of a correct intonation through meaningful phrases and sentences.

Ex.: Asking questions. The student responds to the questions or completes the sentences.

Indirect production of intonation via writing.

Ex.: He taught me how to swim. Indicate the intonation.

Ex.: Similar intonation in two written phrases in the following list: (1) a stone house (2) a nice garden (3) a home

4. Spelling

Ex.: Select the word that is misspelled in each of the following groups of words. (1) auspices (2) cello (3) alternitive (4) proficient (5) none wrong

More complex problems of spelling occur in connection with writing tests.

5. Punctuation and capitalization.

The tasks included in the punctuation and capitalization can be studied by means of similar techniques. The abilities to be learned can be classified as follows:

Use of period
Use of comma
Use of other punctuation marks.
Use of capitals in names
Use of capitals to indicate grammatical relationships
Use of capitals to indicate emotional tone
Other uses of capitals.

Since there are great similarities among different languages in the use of punctuation and capitalization, all the uses described above can be studied within a single test of recognition or production.

Recognition of errors in punctuation and capitalization

Ex.: They climbed into an aeroplane. Soon the aeroplane was high in the sky.

(1) soon (2) Soon

Production of correct punctuation and capitalization

Ex.: Do you plan to come tomorrow ( ) Yes, I do.

(1) (2) . (3) ? (4) :

6. Vocabulary

In the study of vocabulary the words are dealt with as lexical units, not as grammatical units. In the selection of words the similarities between the forms, meanings, and distributions between the native and foreign language should be examined separately.

Recognition of words, oral presentation

Ex.: Meaning of integrity:

(1) intelligence (2) uprightness (3) intrigue (4) weakness

Recognition of words, written presentation

Ex.: He wears glasses (1) for the rain (2) to see with (3) for ice (4) to drink with

Translation as a recognition test

The alternatives are given via the native language.

Production of words

The meaning can be provided in different ways: in foreign language context, in picture context, and in the native language context.

Ex.: What do you see in the picture?

Indirect measures for the production of words

Ex.: I admire him and wish to - banquet as a guest.

(1) a .... t (2) s .... d (3) i . k

7. Grammatical structure

The patterns of arrangement of parts of words into words and the patterns of arrangement of words in sentences. The testing of grammar cannot take place by using similarity comparisons in the same way as in the previous parts of the language ability. The elements of the grammatical ability are grammatical patterns, patterns of words and sentences. An arrangement of words that has a meaning over the separate meanings of the words that constitute it, and is the model for parts of which other words can be substituted without changing the meaning of the arrangement, constitutes a grammatical pattern. The division of grammar into learning tasks is reported in a separate report.

The measures used for an assessment of grammatical ability can be classified as follows:

Knowledge of isolated patterns

Ex.: Write the possessive plural of these words:

(1) woman - (2) donkey -

Recognition of word patterns within sentences

Ex.: (1) Boys (2) The boy (3) I (4) We - strikes the car and runs.

Ex.: (1) Who (2) Whom do you suppose it was?

Recognition of phrases and sentence patterns

Ex.: (1) Although (2) Inspite of being (3) Although I was - barely five, my father sent me to school that fall.

Production of word patterns within sentences

Ex.: I received a letter (write) - in English yesterday.

Partial production of patterns

Ex.: Rearranging parts of a scrambled sentence.

I will send - (1) as soon as (2) you (3) a telegram (4) to London (5) I get

8. Comprehension of meanings in phrases and sentences.

Comprehension consists of all the elements of language which have been examined in the previous sections. It is a more complex task in which an integration of several aspects of recognition is required. Comprehension is examined as a response to both auditory and visual stimuli (listening, reading). The material presented is usually similar except for some tasks: Tables, graphs, etc. material can be included and scanning tasks can be presented in the reading comprehension tests.

In addition to the integration of elementary functions, some other aspects of language achievement can be assessed by means of complex tests. Among them the speed of comprehension (or the speed of stimulus administration) may be of particular importance.

Selection and grouping of ideas to form an overall pattern of thought, skimming of material.

Following directions   Drawing inferences   Reference skills

Organization of parts of a book, reading graphs, alphabetizing, use of index, reading a map, etc.

Some skills which are sometimes included in comprehension were dealt with in the section on vocabulary.

The tests for auditory and reading comprehension are usually constructed according to the following principle:
the student has to answer questions to indicate whether he has comprehended the content played from the tape or the content he has read.

9. Speaking and writing

Auditory stimuli and auditory responses

Ex.: The native speaker asks questions which are recorded on the tape. Immediately after each question another voice from the tape tells the student in his native language what to reply.

Ex.: Description of pictures

Writing by completion

Ex.: The student is expected to complete incomplete sentences

Writing compositions

Style

Ratings based on written compositions

Partial production of written performance

Ex.: Complete the blanks from the context or the translations then check the alternative that fits your answer.

Translation tests are a particular class of writing tests. The production of correct answers is determined by the associations between the native language and English. In these tests the direction of the association chain is a very important factor. Therefore, two types of translation tests should be constructed: one from the native language to English, and one from English to the native language.

Very few examples of the comprehension and writing tests are presented here, because these complex tests are included in all the test batteries of language achievement.

III Some Problems and Procedures Suggested

1. Theoretical approach

From the point of view of learning theory the study of foreign languages provides new problems in addition to these which are common to most school subjects, since the learning situation is different for the students depending on their native language. These problems should be emphasized in the plan for the cross-national study. The most important explanatory concept may be the degree of similarity between the native language and English.

For this purpose accurate comparisons between pairs of languages should be carried out at the pre-test stage by means of both linguistic and educational research. Theoretically it is possible to predict a great deal of the differences in the early achievement in English, if we know how great are the distances between the native language and English in the psychological dimension of similarity. On the other hand, we may not be able to predict in advance what similarities between the pairs of languages are most important, especially at the early stages of learning. Furthermore, the similarity between some structural elements may disturb learning, if the discrimination of small differences is too difficult.

There is, for instance, information available concerning the similarity of the systems of phonemes between various languages. In addition to this information, children can be examined to show what discriminations they are able to make in perception as well as in vocal expression. If these comparisons are carried out at successive stages of learning, very important conclusions might be made.

Scaling techniques should be used at the pre-test stage for this purpose. The number of comparisons will not be too high, if only the similarity between English and each native language is studied. An overall estimation of similarity is not useful. The task should be restricted and defined by selecting some aspects of language for close observation, e.g.,

- phonemes - words -sentence patterns or grammatical structures

Accordingly, the similarity (or the amount of overlap) between English and each native language is rated separately for these aspects, including in the comparison such elements of content and grammar which are usually taught during the first three years of study. Hypotheses should be made on the basis of these similarity scalings.

It is important that the languages to be studied vary greatly in their similarity with English. The list of languages chosen for this part of the study could be approximately the following:

German
Swedish
Flemish (Dutch)
French
Italian or Spanish
Polish or Serbo-Croatian
Hindi
Hebrew
Japanese
Finnish

The final choice may depend on how early English is taught in schools and what are the populations within each country studying English. The main problems can be satisfactorily studied, if at least 6-8 different native languages are included.

The second factor which is essential for any explanation is the frequency of repetition of each learning task. The frequency of administering certain stimuli can be used as an indirect measure of the degree of learning. Many aspects of teaching methods and procedures can be interpreted in terms of the frequency of repetition of some classes of stimuli (or tasks).

This factor cannot be examined very accurately for all the tasks to be learned. Some comparisons are, however, possible. Ex.: Some countries emphasizes a very limited vocabulary in the beginning of learning English so that all the words which occur are repeated frequently. The estimation of redundancy in the administration of stimuli, based on text books etc., provides an index of the frequency of repetition. The direction of associations (native foreign language) may show variations across the countries. Information on teaching methods should give an estimate of the strength of associations in both directions. Emphasis on spoken vs. written language in teaching is also related to the frequency of repetition of various kinds of learning tasks.

In section II (Classification) only such aspects of language ability were examined which occur in common batteries of language achievement. These types of test may not be sufficient in one respect: the amount of generalization of general information or skills should also be studied, because different native languages and different teaching procedures may differ in this respect. Some aspects of adopting general knowledge or skills can be studied by means of 'artifical languages'. The subjects are required to build grammatically correct sentences with some artifical or unknown words.

2. General problems of the cross-cultural study

The cross-cultural comparisons of English as a foreign language should be carried out within the same general frame as the previous part of the I.E.A. project. The background information needed from the countries is approximately the same:

- general values systems adopted in the society
- general educational policies

- educational practices
    - teaching methods
    - curricula, examinations
    - selection and streaming
    - teacher qualifications
- teachers' and students' attitudes (which can be used as both independent and dependent variables in connection with different problems)

On the basis of experiences gained in the previous enterprise some background data should be defined in such a way that the comparability of results will increase.

More detailed information is required in regard to such external factors which may increase/decrease the students' readiness to learn a foreign language:

- additional exercises at school: readers, programmed courses etc.
- second language learned at home
- other languages taught at school (for advanced students)
- the effects of mass media
- general cultural exchange

The teachers' practices should be examined in great detail. It seems to be possible to describe very accurately many aspects of teaching practices in connection with language learning.

3. Analysis of curricula

Some aspects of language ability can be described in very general terms only and therefore the curricula cannot be fixed for each grade level. Ratings concerning the emphasis given to various aspects in English instruction may provide some useful information.

A list of the early vocabulary in English taught at school seems to be important, since the content of items to be presented in problems of grammar, etc., depends on the vocabulary which can be expected to be common to most students. (The average size of vocabulary taught is approximately 300-500 words for the first year, 600-1000 for the second year, and 900-1500 on the average for the third year of instruction.) A list of words which occur in some elementary textbooks has been prepared and comparisons have been carried out concerning the overlap of vocabularies among the textbooks.

For an accurate comparison of the learning of grammatical structures the classification of some standard grammar in English might be used. For less detailed studies a selection of sentence patterns might provide a sufficient background. Many countries do not have any fixed programs in these matters. Lists have been compiled by the subcommittee for both types of comparison.

4. Subjects

If ability in English is examined at an early stage, only the grade level can be chosen as the basis for comparisons (one or two years after the beginning of English studies.) This factor is no longer important at the intermediate or terminal level. Results of a sample consisting of an age group may be less important than in the study of mathematics, since English is a compulsory foreign language in very few countries.

It is suggested by the subcommittee that the hypotheses concerned with the process of learning should be examined for a limited sample of subjects who have studied English one to three years. (Sample A) For an examination of some possible hypotheses it would be important to retest some groups at an interval of one year.

A very detailed test battery should be constructed for Sample A and some tests should be administered as individual tests. Depending on the problems chosen the size of this sample could be about 200-500 subjects in each country.

Another sample (Sample B) corresponding to that used in the previous study should be chosen for the overall description of achievements. No comparisons have so far been carried out concerning the number of English lessons at successive grade levels in various countries. Final suggestions for sampling can be made after such comparisons. It is considered possible that only the Sample 3 of the previous study (pre-university level) should be included for these purposes. For an examination of hypotheses concerning achievement and selectivity, the sampling procedure should, however, be planned in the same way as in the previous study.

5. Test construction

Tests representing the Level 1 of integration (Classification scheme on p.A-11 may be most essential for the part of study which is directly related to learning theory. It is suggested that the following types of test are included for the limited sample (Sample A)

1. Identification and discrimination of sounds
2. Pronunciation, intonation and stress
3. Spelling, punctuation and capitalization
4. Recognition of words
5. Recognition of sentence patterns
6. Active vocabulary
7. Production of grammatical units
8. Auditory comprehension
9. Reading comprehension
10. Speaking
11. Writing

Some of these tests can be combined (e.g. 2 and 10, 3 and 11).

Group tests only can be administered to large samples. Therefore various types of 'indirect measures' of production seem to be necessary for Sample B. It should be examined, however, whether it is possible to collect some material with a direct measure of pronunciation for this Sample B. A written composition test should be included in the battery as a direct measure of production.

The Finnish subcommittee has held two meetings for the discussion of the problems of language testing. The following tasks have been completed:

A general examination of curricula in English in certain countries, as well as overall comparisons of several textbooks. Detailed information for the comparison was obtained especially from Japan and Holland.

An analysis of the vocabulary included in the textbooks or the curricula in the lower grades in some countries. (The tasks was carried out by graduate students.)

An analysis of the sentence patterns taught at lower grades in Finnish secondary schools (Mr. Rauno Piirtola).

An analysis of the structure of grammar taught in Finnish secondary schools at lower grades (Mr. Peter Jarrett and others).

In addition, Mr. Pentti Pitkanen outlined the general classifications of the psychological processes and the content matter, collected the examples, and drew up the theoretical background. Prof. Martti Takala participated in the planning of research problems as well as in the integration of the tasks.

Separate reports were prepared on:

1. List of English words included in the curricula
2. List of sentence patterns included in the curricula
3. List for an analysis of grammar

I. E. A.
POSSIBILITIES OF AN INTERNATIONAL RESEARCH PROJECT IN MOTHER TONGUE AT THE SECONDARY SCHOOL LEVEL
REPORT

As in the research project on mathematics, the preparatory work is organized according to the four following steps:

1) Brief synthesis of study plans in several countries.

2) Examination of knowledge and abilities which it would be possible to test at the following age levels - 13, 15, and 17, i.e., grades 8, 10, and 12.

3) Constructing a table on two axes - one of content and one of objectives in the subject matter teaching.

4) Illustrations of test items corresponding with the cells in the table.

I. Study Plan

Preliminary comments:

1) Except for that part which is concerned with knowledge and techniques general progress in mother tongue is more qualitative and because of this it is not easily analyzable into separate units other than simple indices. The successive programs in the matters follow the following format:

a) Beginning new themes (for example, movement from narrative to dissertation).

b) Movement to successive levels of obstruction (for example, in reading comprehension).

II. In various languages, objective evaluation instruments have been developed independently according to principles which are not identical and often they at different levels. This is the case, for example, with base vocabularies. A preparatory comparative inventory of these instruments is necessary to be able to see whether it will be possible to harmonize the norms.

III. Especially at the secondary level, the use of the mother tongue is intimately connected with intelligent activity on the verbal plane. It is a matter of distinguishing that which belongs to the nuanced use of the language and abilities in judgment and reasoning. Therefore, the application of mental, verbal and non-verbal tests seems to us to be indispensable; moreover, the team in charge of the test construction should have a knowledge of the mental tests chosen in such a way as to avoid confusion.

IV. Even more than in other subject areas, environmental factors, and above all, sociocultural factors have great influence on the nature and level of school attainment. Also, it will be necessary to have an index system defining the various milieux with a fair amount of precision in the student questionnaires. A special subcommittee consisting of proefessors of sociology and education should deal with this work.

V. Certain areas such as literature and spelling are so different from language to language that is hardly likely that it will be possible to apply common tests. Nevertheless, it would be interesting to profit from the occasion of this research to apply identical tests in these areas to samples from countries of the same language. For example, Scotland, England, United States, Australia or on the other hand, French and Belgium.

1. In the objective "Education in mother tongue and by (through) the mother tongue" is often mentioned.

Syllabuses

Documents consulted:

England: G.C.E. papers (ordinary advanced levels)
Scotland: Scottish certificate examination papers (ordinary and higher grade)
France: Time tables and syllabuses of the middle schools (Vuibert, 1961)
Japan: Mimeographed document in English on the objectives of teaching mother tongue in secondary school.
Germany: Syllabuses from Hessen, Hamburg, Bremen and also the Frankfurt tests.
Netherlands: Syllabuses:

The presentation and the structure of these documents vary a great deal; however, all insist on the final important objective-mastering of the language. This is why in the present chapter we will follow the functional hierarchy of the subjects rather than their didactic order. Functionally, mother tongue is the means of communication par excellence. The teaching of it aims at essentially understanding and expression.

1. All programmes insist on the importance of the spoken language; this aspect of study plans will be dealt with first.

2. It is also necessary for pupils to understand the written word of others and to express in writing their own thoughts and feelings.

3. In each language special relatively stable instruments assure social competence in communication; the elements are particularly grammar, spelling and vocabulary. These tools differ considerably in relation to the historical evolution of the cultures, political regimes and the economic conditions of each nation.

4. Certain objectives of a psychological nature can be found in the programmes:

- the wish to see interest in the mother tongue, or certain aspects of it, occupy a high place in the value hierarchy of students;
- the concern for developing taste and literary sensitivity.

13 year old Level

Oral Language

(The syllabus of Hegsen entitles this 'hearing and speaking.')

The European Syllabuses are particularly oriented towards oral summaries of what pupils have read on the one hand and towards recitation or reading aloud of literary works on the other.

The Japanese at this level lay a great deal of emphasis on the spoken language, particularly participation in a discussion on a given theme and training in debating. The Hessian syllabus recommends play-acting.

Written Language

Reading: 1. In continental European countries reading is strictly limited to the study of authors; in particular, literary analysis plays an important part. The main objective is essential aesthetic training. The examples provided draw attention to qualities of style, to the general tone of the piece, to the structure of the composition and to the finer points of detail. Most study plans also recommend reading work either at home or at school followed by a written summary or oral or written commentaries.

The suggested texts differ according to country and language: works from different epochs in France; modern works except for LaFontaine in the French speaking part of Belgium; prose and verse selected from modern literature in Holland; stories; ballads and lyric poems in Hamburg.

2. Independent of literary analysis, Japanese directives lay emphasis on the social use of reading at this level and contact with the thought and emotion of the author; the following objectives are given: adequate use of works of reference, training in summarizing as well as the inculcation of a value: the appreciation of fine literature.

The Belgain (French part) syllabus recommends certain exercises in synthesizing: edit a summary or resume, describe the character of a person, structure of a story.

Composition. The syllabuses are similar:

- in Belgium (French language): accounts, sketches, descriptions;
- in Belgium (Flemish language): descriptions, summaries, letters;
- in France: descriptions, stories;
- in Germany: accounts, summaries, descriptions (views, men, animals, things, machines);
- in Netherlands: essays and letter writing.

The Japanese emphasize certain positive objectives: precision, clarity, cohesivness in work, keeping to the point.

In certain countries, specific learning of essay writing is foreseen:
- in France, establishing a plan and developing a paragraph on a set theme.
- in Belgium (Flemish part), stylistic exercises aimed at looking for the proper term (paronyms, synonyms), division into "alineas"; observation of movements, colors, sounds, etc.

Linguistic Instruments

Vocabulary (certain syllabuses combine this with the study of phraseology).

The Belgian syllabus (French) recommends: word structure, paronyms and antonyms, distinction between literary and figurative meaning (these matters appear two years earlier in the French syllabuses).

The Belgian syllabus (Flemish) also contains the exploration of clusters of ideas on a particular theme (animals, plants, sports, cities, industry . . .) the study of abbreviations in common use.

In Hessen is emphasized the study of words of foreign origin, roots and formation of words, place names, terms in modern vocabulary stemming from the Middle Ages.

Grammar (and analysis). Most syllabuses deal with the nature, form and function of statements as well as the order of words and statements. However, there are certain differences:

-the French syllabus deals specifically with conjunctions, relative pronouns and interrogative pronouns;

-the Belgian syllabus (French) deals with a sense of mood and tone; irregular conjugation and punctuation;

-the Belgian syllabus (Flemish) recommends a detailed study of the nature of words and emphasizes the verb and its conjugation.

Spelling. The difficulties are specific to each language.

The French syllabuses mention spelling and more systematically grammatical spelling. The Dutch syllabuses emphasize the spelling of difficult words and foreign terms.

15 year old Level

Oral Language

The Japanese recommend critical listening to general or literary speeches. Although certain European syllabuses[1] mention audio-visual aids or attendance at "seances litteraires," they are less explicit about the systematic use of teaching methods.

-All syllabuses mention commentary on reading texts, reading aloud and recitation of words which have been subject to literary analysis; some mention practice of dialogue or play acting.

In all syllabuses conversation exercises and discussion on a set topic appear. The Belgian syllabus (French) includes presentation of themes, personnages, authors and epochs followed by an exchange of views. The Japanese suggest initiating pupils to chairing a meeting.

---

[1] Hessen mentions criticism of films, television programmes and plays.

Written Language

Reading comprehension. All syllabuses mention resumes or summaries of texts or works which have been read. As for literary analysis, the choice of works varies a great deal; sometimes they are works from a particular epoch, sometimes from given authors and in some cases only the type of works to study is given.

The Scottish Certificate of Education (Ordinary Grade) emphasizes answering concrete, simple and specific questions on texts which have been read from the fields of drama, poetry, prose and general reading in the first four years of secondary education. Questions on prose should primarily evaluate understanding of the meaning of words and phrases, figurative meanings and the ability to synthesize and evaluate.

The Japanese directives suggest the reading of reports, speeches and criticisms; also the use of reference works, newspapers and magazines. They draw particular attention to questions of developing themes in a story and the intention of the author.

Composition. Syllabuses vary:

-stories, portraits, letter writing, speeches, dialogues, literary or moral themes, planning of arguments (France);
-stories including portraits, amlifications, inter-pupil correspondence, articles for student magazines (French speaking part of Belium);
-summaries, impressions, discussions (Hamburg);
-summaries, impressions, opinions (Hessen);
-letter writing, requests, biographies (Hessen).

The German and Belgian (French part) syllabuses have detailed instructions on stylistics. The French syllabus has precise instructions concerning initiation to the art of writing:

The word and the idea (exaggerations, attenuations, images and comparisons);
The phrase, expression of an idea: the affective phrase; the period; harmony of the phrase and the choice of form; the development and formulation of ideas; study of paragraph construction; intellectual probity; sincerity; agreement of thought and form.

Lingustic instruments

Vocabulary. The French and Hessian syllabuses recommend the occasional study of the history of vocabulary and the evolution of words and their meaning with the "explication de texte."

The Belgian (Flemish) syllabus mentions exercises for improving vocabulary: resistance to Gallicisms, provincialisms; improvement of texts by introducing variety and precision. The Hamburg syllabus suggests word formation and regional forms of vocabulary and language.

Grammar: Most syllabuses suggest the learning of grammatical notions when dealing with "explications d'anteurs" and the correction of essays. The French-speaking part of Belgium maintain that these are insufficient and that formal instruction in grammar should be continued at this level; it should consist of the revision of difficulties of agreement and conjugation; in syntax--word order, inversion, outline in harmony with the intention of the style. The Japanese emphasize an understanding of grammar rules.

Spelling. In general, this is only taught very occasionally at this level. However, the Belgian (French part) syllabus lays down the continuation of dictations at this level.

17 year old Level

Note: The time table of terminal classes in France is cut down by an hour a week in Mother tongue; the syllabus is mostly literary and is left to the initiative of each teacher.

Oral Language

-Presentation and explanation of literary works, development of a theme (French speaking part of Belgium).
-Expositions in conjunction with essay writing (Flemish speaking part of Belgium).
-Speeches, spoken criticism of films, television programmes and plays (Hessen).

The Japanese emphasize the playing of a useful role in meetings and they put much stress on students developing the ability to defend their opinions clearly (affirm, explain, prove).

Written Language

Reading. This is more or less restricted to the explanation of literary works; critical, historic, philosophical in conjunction with courses in the history of literature and "literary aesthetics".

In the Netherlands, mentioning the reading of literary works from various periods, mentions in particular, making the students aware of the period when they were born.

In Germany special mention is made of translation of foreign authors: in Frence (16 years) the reading of ancient authors is recommended in the "modern languages" track (section moderne).

In Japan the following objectives are outlined: formation of personal and critical thought; a deepening of the knowledge of the language; development of senstitivity. They suggest dealing with logical and philosophical points of view within the area of literary appreciation. They also mention literary criticism, the reading of various articles culled from reviews.

Composition

Belgian syllabuses, dissertations, critical analysis of controversial texts, summaries of speeches, free work (French); dissertations written portraits of statutes, letter writing, syntheses (Flemish); souvenirs, interpretations, summaries of discussions (Hamburg).

In Japan the long term objective of this teaching is to develop logical thought and creativity.

The Scottish instructions[1] define the requirements of success in exams as "a sufficient ability in composition and clear exposition in grammatically correct English."

The ability to tackle more broader topics in successive classes in secondary education is stressed in various courses. Hamburg makes this quite precise in terms of length of work: from 1 hour at the 13-year-old level to 3-5 hours at the 17-year-old-level. In Oliver's work on the Joint Matriculation Board it is the ability to tackle different domaines which is evaluated: the pupil must develop shortly (about 1/2 hour per topic) 3 themes to be chosen from, on the one hand, aesthetic questions and political, economic and Philosophical questions on the other hand.

Instruments

Vocabulary. This is only dealt with very occasionally at this level. However, terminal examinations in Great Britain do take account of this aspect of mother tongue either in separate tests (Oliver) or in reading comprehension questions (Scottish Certificate of Education).

Grammar: In Belgian (French part) the study of questions of syntax grammatical commentary on some mode in texts, and the discussion of language problems are suggested.

In Flemish (Belgium) teaching at this level is more oriented towards sociological and historical aspects of forms of language.

In Holland a broad study of grammatical questions is advised.

In other syllabuses allusion is made more to this is "explication de textes litteraries".

Spelling. Only the Belgian (French part) syllabus mentions the testing of spelling by use of dictations.

Domains to be Explored

I. Spoken language.

The objective examination of the spoken language does not seem possible to us within the framework of an international research project.

---

[1] See page 8 in French version.

Even when limiting ourselves to countries of the same language and taking special sub-samples there are still problems which can only be overcome with difficulty: complete equality of physical conditions when hearing the sounds, establishing acceptable criteria which overcome the diversity of local accents and a sure method of analysis. These are beyond the present resources of educational research.

An attempt to measure spoken language understanding would be interesting and would be within our means on condition that in each country the television authorities were willing to help by showing the same film over the whole country. The subject should, of course, be stimulating and capable of comprehension at all levels (this latter point should be the object of a preliminary study and of the pre-testing of the broadcast).

The test should be similar to a silent reading test; questions on the words spoken, appreciation of the language of the characters and their behavior through their language; items attempting to evaluate an understanding of the whole or an important part of the sequence.

As for the progressive difficulty of items, it should be possible to use the same film at all three levels; this should considerably diminish the work involved.

II. Written Language

A. Reading Comprehension

This is the domaine where international research is easiest and where the terms in the 'Taxonomy of Educational Objectives' apply most completely. There are, however, some special problems to be resolved.

1. It should be possible to present texts where the main effort of understanding would have a literary bias.

2. The choice of passages raises the tricky matter of texts equivalent in interest and difficulty for all the national samples; preliminary work could be undertaken on this by a small committee whose task would be to submit suggestions to N.C. and get their reactions to them.

3. At the secondary level, the problem exists of the distinction between a reading comprehension test and a verbal intelligence test. To what extent does compulsory reference to the text constitute a sufficient criterion? By using a particular method of statistical analysis will it be possible to identify these two factors?

4. Below are some suggestions concerning what might be dealt with at the various levels:

Level 1 - 13 years

a) Ability to find a particular idea in the text.
b) Ability to find positive arguments in the text.
c) Ability to find various types of motor or sensory details in the text.

d) Movement from a text to a symbolic or graphic representation.
e) Movement from a literal to non-literal sense or vice-versa.
f) Adequate use of reference works.
g) Knowledge of current abbreviations.

Specially for literary texts:

a) Outline particular qualities of style.
b) Outline the characteristics of certain details
c) Description of the psychology of a character.

Several delicate problems will, however, have to be first of all, resolved.

The proposed theme should be equally stimulating for all "milieux," whether it be the course of study (programme), rural or urban character of the regions or the language of the country. As with reading comprehension, a sub-committee could work out a list of themes which could be submitted in each country for judgment by n teachers in schools not in the sampled schools according to a rating scale technique.

If all the subjects were kept, the work load would be prohibitive, but it should be possible to bring the work down to an acceptable size by limiting this section to a fraction of all the essays.

At each level, several types of themes are dealt with; since ability in essay writing is not a simply entity (a pupil's performances in story-writing, letter writing and dissertations are far from uniform) it would seem desirable to have two subjects with a short time allowed for each rather than only one subject with a considerable time allowed for it.

A global assessment of each essay would be of little interest and would induce very questionable comparisons. On the other hand, useful information may be gained from a standardized analtyic notation with total means at zero--an analogous system to that used in the pilot study; the comparison of profiles (or models with n dimensions) thus obtained would render meaningful information; in particular there would be profile patterns of each linguistic or national group.

Furthermore, there would be meaningful information about differentiated sections (e.g., sections d'humanities anciennes et modernes) within a national system of education.

It would be a matter of choosing those tasks appropriate to what is imposed at each level (descriptions, narrations, dissertations, essays . . .); for each of these sub-divisions it is possible to establish a 3, 5, or 7 point rating scale; previous studies have indicated methods of quantifying these types of criteria. The reliability of such scoring is satisfactory if the same marker is kept to one question and if his consistency is checked by the retest of a batch of papers.

This undertaking would not be an isolated case; in the literature are a certain number of researchers which may be of use to a work group; for example:

- appreciation of ideas (clarity and continuity of thought, originality of ideas, their interest, vocabulary); appreciation of structure (variety of phrases, correction of their structure) (Schonell, 1942);

- depth (clarity and continuity of thought, originality of ideas, form (complexity, variety, correction of form, wealth and precision in vocabulary) (Institut Superieur de Pedagogie du Hainaut, 1947);

- analysis of movement, modes of sensation, imagery, emotion, ideas associated with typical profiles (centre de reserche sur l'expression francaise (3);

- readability; Flesch's index - De Landsheere (4).

Some types of possible topics, appropriate to particular programmes.

Level 1-13 years
  a) A short description
  b) A brief account of a personal experience
  c) Summary

Level 2-15 years
  a) Narration
  b) Portrait
  c) Letter writing
  d) Intervention in a debate
  e) Dialogue
  f) Article or survey for a student magazine.

Level 3-17 years
  a) Dissertation (long essay)
  b) Lecture
  c) Summary of a lecture
  d) Critical examination of two opposing points of view.

Note: Other tasks in this context may be suggested; here are some examples:

- The French suggest for the 13 year olds the following: for a particular essay title, work out a plan.

- Complete a sentence freely where the first few words have been given. Binet and Marg. Evard used this sort of test in trying to identify certain types of inspiration: numerical, observational, imaginatory, affective and emotive.

Certain authors have used this type of question as test items awarding one point, a half point, or zero according to given criteria on the quality of the answer. We are not in favour of using this sort of test: on one hand it is difficult to get reliability internationally; on the other hand, the existence of model solutions means that original answers are penalized.

III. Instruments

A. Vocabulary

Many types of tests have been used to try to assess pupils' knowledge of this domain of mother tongue. In order to differentiate between recall and recognition, certain authors have used both open-ended and multiple choice items; despite the rich documentation on the first type, we suggest that it is not worth consulting it because of the numerous weaknesses there would be in this context in a large international study: length of time required for learning and particularly between-language equivalence in the selection of acceptable synonyms.

The types of possible tests are numerous, for example:

- synonyms and antonyms.
- choice of work (from among others) to cover a common general idea.
- abstract word covering several related concrete words.
- choice of word best completing a given phrase.
- choice of one word in a series expressing the same idea, only to different degrees.
- from a given group of prefixes or suffixes, choose the one which has the common root.

In an international study the delicate problem is that of constructing batteries of equal difficulty in all countries. Well-known words in one language may be little known in another; thus 'parterre' was one of the most difficult in the states on Termann-Merrill's vocabulary scale but was on the 7th rung in the French language. To avoid this it would be necessary first of all to construct base vocabularies and then select items from the same frequency or difficulty zones in all language groups.

Level 1-13 years

a) Knowledge of general rules of composition.
b) Use of terms (semantemes,pronouns, morphames)
c) Ability to graduate degrees of same idea
d) First distinction between literal meaning and figurative meaning

Level 2-15 years

a) Recognition of nuance of meaning.
b) Recognition of nuance of degree.
c) Old words and expressions.
d) Neologisms and their classifical equivalents.

Level 3-18 years.

Some of the more difficult preceding items.

B. Grammar

Most domains in grammar are specific to each language. However, in some aspects it is possible to ask common questions.

a) Items appropriate to the nature and function of words and prepositions.
b) It might be possible to examine the operation of certain rules of punctuation on an international scale; under what conditions would this be possible.
c) It would be possible to have common tests in certain aspects of phraseology--making a sentence more concise, putting into direct or indirect language.

But syntax, time, voice, moods of verbs and agreements can only be dealt with within one language group.

With this in mind, here are some suggestions for items on an international scale.

Level I-13 years

a) Nature and function of words and propositions.
b) Rules of punctuation.
c) Putting into direct or indirect language.
d) Putting into a more concise form (e.g., form one sentence having the same meaning as two other given sentences; eliminate in a text the repetition of a noun or verb.

Level II-15 years

a) Correct use of prepositions, subordinate conjunctions and relative and interrogative pronouns.
b) More difficult items from level 7.

Level III-17 years.

The same questions as at Level 2 in terms of measuring the progress made during the last two years.

C. Spelling

In grammatical spelling, the rules vary completely from one language to another and thus a common test would be impossible.

If spelling scales existed according to the same norms in the different language countries in the research project, a comparative study would be possible starting with tests consisting of words from the same difficulty zones. But these tests probably do not exist at present and the construction of them would take several years of a research study to itself.

D. Affective domain - interests and attitudes.

In the second volume of their "Taxonomy of Educational Objectives" Bloom and his Associates (1) have classified the general affective aims in education according to their intensity. They distinguish the following progressive steps:

- reception
- active response
- the evaluating of these things in the minds of the pupils
- the integration of these values into a hierarchial system
- the influence of this system on behavior or philosophy in life

It would be interesting to examine what place the various aspects of mother tongue have in the hierarchy of pupils interests and attitudes. For sure, the results depend to a large extent on the quality of the teachers but the great number of schools will cancel out individual differences and would give a reliable enough picture of the whole.

Various means could be used towards this end:

a) Preference choice questionnaires between the various aspects studied and also subdivisions thereof.

b) A questionnaire analogous in form to Allport and Vernon's Study of Values.

c) A questionnaire adapted from vocational choice.

d) An interest test in the form of general knowledge test: from among a certain number of world known men, indicate what has made them famous.

e) On the pretext of editing a journal ask the student to choose from x items three successive choices which he would like to keep and three others which he would not like to have.

1) Krathwohl, D. R.; Bloom, B. S.; Masia, B.B. Taxonomy of Educational Objectives II Affective Domain (New York: David McKay, 1964).

Level I

		A Knowledge	B Comprehension	C Analysis	D Synthesis	E Appreciation	F Interest	G Literary Sensitivity	H Creativity
Spoken Language	1. Comprehension								
Written Language	2. Comprehension functional aspect								
	3. Comprehension Literary aspect								
	4. Expression								
Tools	5. Vocabulary								
	6. Grammar and Analysis								
	7. Spelling								

Level II

	Spoken Language	Written Language			Tools		
	1. Comprehension	2. Comprehension functional aspect	3. Comprehension Literary aspect	4. Expression	5. Vocabulary	6. Grammar and Analysis	7. Spelling
H Creativity							
G Literary Sensitivity							
F Interest							
E Appreciation							
D Synthesis							
C Analysis							
B Comprehension							
A Knowledge							

Level III

	Spoken Language	Written Language			Tools		
	1. Comprehension	2. Comprehension Functional aspect	3. Comprehension Literary aspect	4. Expression	5. Vocabulary	6. Grammar and Analysis	7. Spelling
H Creativity							
G Literary Sensitivity							
F Interest							
E Appreciation							
D Synthesis							
C Analysis							
B Comprehension							
A Knowledge							

Conclusions

International research on outcomes in "Mother Tongue" at the 13, 15 and 17 year levels is not impossible even if it is not complete.

It will hardly be possible to begin to consider (in a comparative way) those aspects which differ totally according to the language, in particular literary history and literature, spelling and certain aspects of grammar.

But work is possible in reading comprehension (and to a less extent comprehension of the spoken language - or oral comprehension) composition, analysis, vocabulary (if comparable base vocabularies are used).

There is also a unique opportunity to examine other aspects of language through common tests for countries using the source language. It would also be opportune to try to estimate, at least at the terminal level, attitudes of students towards the Mother Tongue itself and different aspects of it.

The work should be different from that in the mathematics phase in that the orientation of the research should be even more descriptive than normative: it is profiles or standard test models which should be compared rather than raw scores.

The variety of possible types of tests is considerable in each domain. It should be possible to limit ourselves to 2 or 3 forms but the tests should be so graduated as to allow a comparison of progress from one age to the next and, at the same level, from one track to another. The total tests consisting of 4 main blocks could be administered as follows: 13 year old level: A and B; 15 year old level: B and C; 17 year old level: C and D.

As a result of our considerations, we think that the preparatory phase of such a research project would take about 3 years. The next phase of IEA should therefore be devoted to another subject area; but, in the meantime a committee should be set up consisting of specialists in teaching and testing; this committee would then decide on the various suggested options presented in this report after consultation with the National Centers. At the same time various sub-committees would carry out an analagous work for countries using the source language.

This undertaking would be long and laborious, but its importance is such that the effort would be well worthwhile.

- - - -

The committee consisted of:

Madame Bowijts - Secretary of the University Consultative Committee on Education
Mr. De Landsheere - Chef de Travaux (Supervisor) at Liege University
Mr. Jeanfils - Teacher at Malimody - In charge of Mother Tongue research
Mr. Rigaux - Lecturer at Charles Buls Training College, Brussels
Mr. Y. Roger - Secretary General of Middle School Dept., Ministry of Education and Culture
Mr. F. Hotyat - Reporter

Proposal for an International Study of Student Responses to Literature

July 28, 1965

Arthur W. Foshay, Teachers College
Columbia University
Alan C. Purves, Educational Testing
Service, Princeton,
N.J.

I. THEORY OF THE PROJECT

A. The Relation of the Study to Educational Problems

We propose to describe how students and teachers interpret and evaluate a work of literature, and the interaction between them. The study we anticipate here will seek to discern what intellectual operations, and patterns of operations, students and teachers use. We seek not to measure competence within these patterns, but simply to describe their nature and order; not how well students do, but what they do and the order in which they do it. Nor do we seek to measure taste. We seek to describe the premises of taste. Later, we propose to describe the preferences for various types of literary styles and content, and also to compare the preferences of students and teachers. All of this we propose doing on an international scale.

We propose to examine these national patterns (for we suppose that the patterns will be national in character) among students of differing age levels, intellectual abilities, sex, and (where appropriate) educational experience, ethnic background, and geographical region. In addition, we expect to investigate the patterns of preference of teachers in the light of their teaching experience, and amount and kind of training.

In proposing this study, we seek to deal with two interrelated problem-areas: the school and schooling on the one hand, certain aspects of society on the other. With respect to schools, we hope that what follows will make apparent the kind of pedagogical problem we hope to clarify through this study. With respect to society and culture, some comments will appear in later pages.

In conducting both of these descriptions, we seek to define the nature of the gap between the "literary establishment" and the popular mind. That such a gap exists has been a long-standing observation; the existance of the gap has greatly complicated both certain aspects of public decision making and the development of sound pedagogical practices. The teacher has not known whether to impose a set of values, or to work with the base that exists in the student's world. In the United States, this "schizophrenia" has resulted in the muddled compromise of our textbooks which juxtaposes Shakespeare and TV Westerns, treating both with equal seriousness and essentially the same intellectual equipment. There is, moreover, a gap between the expressed goals of literary critics and of teachers of literature, a gap in taste, in objectives, and in tasks performed. Such gaps, if left latent and unexamined, eventually produce strains of misunderstanding among groups of the population that lead to difficulties in the making of large-scale decisions, and even, upon occasion, to a kind of social disaffiliation.

One can describe these gaps by juxtaposing the preferences of the three groups--students, teachers, and disciplined critics--and showing the nature and degree of the incongruity among them.

With respect to educational planning, this study can, by showing gaps, indicate possible bridges. It can show the teacher both the modes of thinking preferred by a literary elite, and those preferred by the group he seeks to educate. If the teacher is to be the bridge connecting the elite world and the popular world, he must of course have means for knowing both worlds equally well. As things stand now, satisfactory means for knowing how students think about literature are not available.

B. Relation of critical attitudes and attitudes toward life

The need for drawing a picture of today's critical temper is important in itself. But significant literature is an author's rendering of life. It follows, therefore, that the way in which a person is taught to approach literature may well be reflected in his approach to life. Strength and subtlety in one may well parallel strength and subtlety in the other. The intellectual historian can assess the climate of opinion of an era by examining its art and its critical method and standards. For example, the eighteenth-century standard of decorum was applied not simply to the work of Boileau, Racine, or Pope, but generally to habits of mind, and one may infer the habits of mind of the era from the study of these authors. This standard influenced attitudes towards religion, government, psychology, warfare, and social behavior. Pope's Essay on Man is an aesthetic standard made a code of life. The reflection of attitudes toward life in attitudes toward art and critical postures remains valid in age after age (It is paralleled by the reflection of attitudes toward life in attitudes toward rhetoric and style, and can form the basis for a similar study of rhetorical preferences). The assumption about the relationship of art and life underlies the whole field of intellectual history, which often generalizes about the values and outlook of a society from its view of art and criticism, and, more, from the view of literature held by the critic, teacher, or student and the world from which it arises.

Since literature and criticism are taught, we can see in them at least the covert teaching of a view of man, as well as the student's reaction. We can see, for instance, the conflict of forces that occurs within a student nurtured on a sub-literature--television, radio, and popular story magazines--who is exposed to a highly disciplined study of great works. That a tension exists between the popular culture and the great culture, we know; perhaps we can come to understand something of its nature. In different countries, of course, this tension exists to a greater or lesser extent, but it is the extent which we, in part, seek to determine.

C. International implications

The third goal is, of course, the comparative one, for as the modes of criticism have differed between eras, so have they differed between countries in the same era (as a look at published German and English criticism of the late 19th Century would immediately show). One would suspect that although a comparative look at critical preferences might turn up not a single preferred mode but several, the pattern of these preferences might differ from nation to nation.

The exposition of these national patterns will serve the most important function of increasing a national educational system's awareness of similarities and differences between nations. We suspect that cliches and popular conceptions contain a modicum of truth: there is, in some meaningful sense, a characteristic American, or Englishman, or Frenchman, although there are strong lines of cultural unity among these. It is these characterological differences, just as much as linguistic differences, which hamper international understanding and communication. These differences are in part differences in ways of perceiving. One may take each experience, like each literary work, as a new one and seek to treat it in its own terms. One may, on the other hand, look at each work or each experience through a categorical lens, see it in terms of tradition, or prescribed classifications. There are analogous differences in ways of interpreting and evaluating a literary experience, which parallel ways of interpreting and evaluating any experience. Although these differences may be subtle and hard to discern, they are vast in import. This study will, in part, depict both differences and similarities. Such a depiction will itself be valuable because it will make a particular teacher or nation aware of the existence of other modes of perception, interpretation, and evaluation. It is the fostering of this sort of awareness that is most important for a group such as IEA.

II. BACKGROUND

In making the preliminary investigation, the first step was to write a number of critics--men eminent for having discussed the varieties of criticism--and from their replies to the question "What are the accepted modes of criticism?" to arrive at a working summary of current critical modes. To the same end, we also examined a sample of college-freshmen and high school essays on a given story, and a sample of teacher responses to the question, "How would you teach this story?"

From these materials and from a subsequent conference of scholars, we developed a tentative classification of critical elements (see the attached documentary prevues of the elements of criticism and interpretation). The scheme, arbitrary as it is, accounts for most of the intellectual operations of the critical act, and with a slight amount of refining can become an operational tool. Yet, how to determine what critical apparatus obtains in the work of a given student or teacher is a knotty problem, for there are certain imponderables that need to be accounted for.

The first is that different literary genres or individual works may call for differing critical approaches and might exclude the student's habitual approach. As our preliminary study showed, one work we selected called for a symbolic reading, despite the resistance a student or teacher might have for the superimposition of typological structures; another called for a social interpretation, again despite resistance.

The second is that a certain tension exists between the private and public reaction to a work of literature. We might say, the modes of discourse about a work differ depending on the critic's audience. We must consider the very possibility of determining a "true" response. It often seems that there is only a rhetorical connection between the parts of a student's response; for instance, his interpretation does not logically follow from his analysis.

The third is that since this is a preferential test, the choice must be free. We have found that when we offered a list of alternative interpretations (orthodox ones by our standards), the responses fell into a "taught" and a definable pattern, but that when the students had to make their own interpretation, a majority of them gave a reading not included in our list--in fact totally at variance with what their peers chose. It fell outside of our hypothesis, yet was obviously predominant in the minds of students, and, presumably, of teachers. Even were we to include more alternatives, we might miss the initial response of the reader. Further, we might present him with one that had not occurred to him, that would not have followed from his pattern of analysis, but that would have charm.

The final imponderable is how to escape bias as among schools of criticism in interpreting the data. We trust that the terms used in the classification are neutral, as well as exhaustive.

III. RELATED RESEARCH

One may truthfully say that there has been little major research in the description of student criticism in the past thirty years. Over all studies looms the influential Practical Criticism of I.A. Richards, the first attempt to examine objectively the writing of students about literature. Richard's work, important to modern criticism as it is, deals with a rather limited group (Cambridge undergraduates), and is in no way comparative. The technique, however, the presentation of an unknown work with a question leading to a free response, cannot be improved upon, and is, in fact the basis for this study.

A similar study, Experiences d'Analyse textuelle en vue de l'explication litteraire: travaux d'eleves by Servais Etienne (Paris, 1935), demonstrates selected students' progress in the genre of the formal explication de texte and comments on their successive attempts to relate form to content in longer, prepared compositions. In restricting his study to a small group of students (aged 17 to 19), M. Etienne is able to compare the degrees of proficiency attained by some of them. He is, however, chiefly concerned with the excellence rather than with the nature of his students' writings.

Related research in the varieties of modern critical approach is also dominated by one work, Theory of Literature by Wallek and Austin Warren, which comprised a survey of current critical method. The book has been imitated and, to some extent, brought up to date in many histories of contemporary criticism and aesthetics, such as those of W.J. Bate, Rene Wellek, R.S. Crane, et.al., but none of these, naturally enough, go beyond the cream of published criticism.

In the field, research, particularly in England and the United States, has been devoted primarily to composition, as Research in Written Composition, by Braddock and Lloyd-Jones has shown. Very little of this research has dealt with the ways in which students react to works of literature, nor has much of that research that has been devoted to reading, for the latter has been devoted to the cognitive problem of reading. The only study of critical reading that comes to mind is John Carroll's "Vectors of Prose Style," which bears a minimal relation to the ends of this study. Some good articles have emerged from the study of an individual class or an individual teacher's work, some good objective tests, such as those of the University of Chicago, and the Carnegie Institute of Technology, have evolved; little of this bears on this study, because it emanates from a critical bias, and it is designed to see whether students have learnt a particular task well, not to find out what tasks they have learnt.

IV. HYPOTHESES OF THE PROJECT

A. Definitions

1. A sentence is a (grammatically independent) group of words which expresses a distinguishable thought.

2. A leading sentence is a sentence that expresses an independent whole idea.

3. A topic sentence is a sentence that expresses an idea contributed to by one or more other sentences.

4. A supporting sentence is a sentence which serves to illustrate or expand upon a topic sentence.

5. An element of criticism or interpretation is a specific act of engagement, perception, interpretation, or evaluation of a work of literature.

6. A category is a grouping of elements. In this study there are four categories: engagement, perception, interpretation, and evaluation.

7. A pattern of criticism consists of a configuration of elements irrespective of category.

8. Frequency of an element or category is the percentage of students using that element or category in a topic or leading sentence, regardless of the number of sentences devoted to that category (whether a student mentions plot once or six times is immaterial).

9. Intensity of an element or category is the average number of sentences devoted to that element or category by a particular population of students.

B. Assumptions Underlying the Project

1. A pattern of criticism is manifest in a written comment on a specific literary work.

2. A pattern of criticism may be discerned by analysis of a written comment on a literary work.

3. Patterns of criticism are partial indicia of patterns of reacting to, perceiving, interpreting, and evaluating experience, or of attitudes toward the world (see Section I of this proposal).

4. The definition of elements identifies the major approaches to a work of literature in the countries of this study.

5. The patterns of criticism will not be identical with the patterns of rhetorical composition (This assumption may be proved by a parallel study of rhetoric).

C. Hypotheses about Nations

1. Each nation can be identified by prevailing patterns of criticism.

2. The number of dominant patterns that exist for different countries will vary.

D. Hypotheses about Students

1. The student's pattern of criticism will be a function of the educational system he is in.

2. The student's pattern of criticism within a system will be a function of:

    a) age
    b) sex
    c) educational experience
    d) socio-economic status
    e) kind of school
    f) region
    g) teacher training

3. The student's pattern of criticism will be related to his stated literary preferences.

4. The student's pattern of criticism will be related to that of his teacher and to the "official" pattern of established scholars and critics.

5. The array of patterns will broaden as the student becomes more advanced.

6. Students will have a smaller array of patterns than teachers and teachers than critics.

E. Hypotheses about Teachers

1. The teacher's pattern of criticism will be a function of:

    a) training (highest degree or certificates held and principal field of study)
    b) stated reading preferences
    c) length of service
    d) sex
    e) educational and socio-economic status of students
    f) size of school
    g) rural-urban status of school
    h) perception of student's literary potential

F. Operating Assumptions

1. A sentence may be categorized according to the element it exhibits.

2. A topic sentence on one element may be supported by sentences on another element. (A sentence on character analysis may be supported by one or more sentences on narrative line, on diction, or on syntactic analysis.)

G. Operational Hypotheses

1. The frequency of a given element will differ by country.
2. The intensity of a given element will differ by country.
3. Contingent relationships among elements will exist by country.
4. For a given element in a topic sentence, the elements employed in supporting sentences will be the same from country to country (e.g., students, regardless of country, might support a character analysis by reference to narrative rather than by reference to external moral criteria.).
5. The average number of topic sentences in an essay will not differ by country.
6. The average number of leading sentences in an essay will not differ by country.
7. In a country, the frequencies of a given element expressed in leading sentences and topic sentences will not differ.
8. The frequency of a given category will differ by country.
9. Specific associations of categories will differ by country.
10. The intensity of a given category will differ by country.
11. The number of categories employed by students will differ by country.

V. PROCEDURE

Since we seek to describe critical responses, we will always keep in mind three intentions: the study will not evaluate taste; it will not follow a particular critical line; it will not compare achievement.

With these in mind, we may consider the following questions, each of which names a successive phase of the project:

1. To what extent are responses to literary works classifiable?
2. To what extent are preferences for particular responses reliable?
3. What are the modal patterns of response among students and teachers in different countries?
4. In what way can the response preferences be characterized as literary responses and as types of Weltanschauungen?

The present summary and scoring sheet constitute the first step toward developing a system by which we may clarify responses. The questions we are now asking of it are:

1. Is it exhaustive in terms of genres, in terms of response media, and in terms of population?
2. Are the categories mutually exclusive?
3. Is it reliable, in that it will function identically with varying texts and readers?

The answers to these questions have begun to be determined first by polling a larger population of established critics, both in the United States and particularly in the other countries involved in the study and asking them whether they can think of any additions to the summary. The next step was to use the classification in making an analysis of a variety of responses, and the third was to use it with a number of readers to determine inter-reader reliability. We further intend to test the system with responses to poetry and prose, with oral and written responses, and with both student and teacher populations as advanced and slow classes, highly trained and untrained teachers, and students of different national origins. This pilot study has been accomplished with a small group of students and teachers in the New York area, and with other small groups from Belgium and Germany.

A. Rationale

The rationale of these pilot investigations is as it will be for most of the project. The students are offered a short work (3-5 pages if a story, under 30 lines if a poem) and are asked to read and "discuss" the work. This question enables them to frame a response most suitable to them, and enables us to compare that response to the expressed teaching preference of the instructor.

The essay is then read by three readers to assess inter-reader reliability, and to assure accuracy, each reader using a scoring sheet which shows the element and type of each sentence that a student writes. The scores will then be analyzed for significant patterns in accordance with the operational hypotheses above.

A similar form of reading and analysis will be used on the response of teachers to the question, "How would you teach this work to your class?"

The second step is to develop a "parsimonious" instrument. We must determine:

a) Whether the responses of a student are reliable within a genre and between genres.
b) The extent to which a particular literary work determines a particular response pattern.
c) Whether the responses are reliable across oral, written, and multiple-choice media.

To make these determinations, we propose to present a student population and a teacher population with two works of poetry and two of prose, and ask them to respond to one in each genre orally and to one in each genre in an essay. Using the classification, we will make a content analysis of the responses and determine the reliability of the responses. If they prove reliable (which, grossly measured, we expect they will), we can then explore the reliability of a more easily scored instrument, checking the responses to this against another group of essays and interviews. If this sort of instrument proves unreliable, we will seek to determine the least amount of "free information we will need to determine a response preference. Even with an easily scored instrument, we suspect that at every stage of testing we will need a modicum of free response material to serve as a check and as the source of finer distinctions.

For the taste-preference section, we will assemble a variety of short tests--typical paragraphs from sub-literature, from best-sellers, from avant garde literature, from academic literature, from "philosophic" literature, to name but a few of the array--and present them to teachers and students. We will ask them to rank their preferences (both what they like and what they think good by stylistic, moral, affective or whatever criteria) and see what sort of comparability exists between the responses in a population and across two populations.

During the course of this phase, we will also be continually testing the literary works to see whether they pre-ordain a certain response. Further, we will seek to determine the sort of literary work best suited for a valid response. In making these determinations, we will move to a larger population and seek to determine comparability across sub-groups-- ethnic, cultural, intellectual, sex, grade-level, school--and between teacher and student responses. From this we will develop a test and analysis procedure that can be operational and efficient on an international level.

B. Population

The population for all phases of this project will consist of thirteen year olds and seventeen year olds. The first group is the generally comparable one of these children in their last year of elementary education; the second that of those who are in the last year of secondary education. The sample will, of course, be such as to enable analysis according to Hypotheses D and E (above).

C. Instrumentation

1. Critical patterns instrument (two forms)

    a) Derived from content analysis, it will emerge in a machine-scoreable form, but will be supplemented by a small number of free responses.

2. Preferred literature instrument (rank order of passages by varying criteria).

3. Inventory of student's attitudes towards the field of literature.

4. Inventory of teacher's attitudes towards students.

5. Background questionnaire concerning teacher and school.

6. Background questionnaire concerning students' and teachers' educational and home experience of literature.

As in an exploratory study we have completed, we will present more than one work to each student and teacher, and in some cases, more than one test. This will again be to determine the reliability of the response and provide a clearer view of the patterns.

Because of the problems of literary translation, and because of the cultural differences that might intervene in the response to one literary work (the inability of a student from one country to appreciate the conflict in a work from a radically different society) we will attempt to devise a test that can use an indigenous work but apply to it a set of neutral questions; yet we suspect that a better solution would be to present a "neutral" work. Then, if a particular nation wishes, it may use the form of the test with an indigenous work to arrive at a more accurate assessment of its own sub-groups.

D. Analysis

From the data gathered, we will determine the modal response patterns of a given group or sub-group. We will also determine the interrelation of modal response patterns of students and teachers in a given educational system.

The final phase of the project will be the characterization of each of the major response patterns. We will do so by returning to the classification and inducing from each major pattern its dominant features. These features will be examined to see if they are interdependent and in what ways. From this characterization, we will proceed to define the critical attitude it implies and to make a tentative inference as to the Weltanschauung it implies.

The characterization will proceed to relate the response patterns to the variables gathered in the test (sex, age, etc), but more importantly perhaps, to those variables which given some indication of the reading or amusement habits and preferences of the various population. From this relation, we can infer the interrelation (or lack of commonality) between popular "literature" and prescribed literature, between student and imposed literary values. This inference will be somewhat tentative, because there are obviously a greater number of variables acting on any one person than one can ascertain or even suppose.

The inferences about Weltanschauungen will arise from a consideration of the implications of each pattern. If, for instance, one pattern shows a highly formal system of analysis and no evaluation, a possible inference would be that a "scientific" attitude obtains in all fields of inquiry--an attitude which says that the world exists to be examined but not to be judged, further that man is a logical, but perhaps not an evaluating instrument, and that the world exists as objective phenomena not to be viewed subjectively. Contrasted to this view, might be that presented in a pattern of moral interpretation and evaluation with perception limited to the search for allegory. Such a pattern would indicate a typological or even anagogical man, one who sees things in symbolic terms, who sees all reality in terms of paradigms of behavior, as a series of dimunitions from one central standard. Obviously, these two views cannot be compared in any evaluative fashion, but they can be pointed out.

We have said that these inferences are tentative, and so they are. They are suggestive generalizations based on the ways in which students and teacher perceive, describe, and evaluate a given literary object.

E. Time Schedule

The perfecting of the instruments would require one year, in which time the preliminary data should yield instruments 1 and 2 with sufficient cross-reference material derived from essays and interviews to insure its reliability. The other instruments can be supplied by that time.

Testing, scoring, and analysis should require two additional years with additional time for publication of the results.

VI. PERSONNEL

The director of the proposed project would be Prof. Arthur W. Foshay, Associate Dean of Research and Field Services, Teachers College, Columbia University. The principal technical consultant during the developmental phase of the project has been Prof. Alan C. Purves, who has left his post as Professor of English at Barnard College, Columbia University, to become Professional Associate, Humanities Department, Educational Testing Service. It is expected that Dr. Purves will continue to consult with the project, which will also continue to have access to the other professional services of ETS.

In addition to the director and general consultant, the project will require consultants in literature from several of the participating countries, and also a staff of technical and clerical assistants.

We assume that the administrative arrangement used in the I.E.A. Mathematics study would continue to serve this project: i.e., a council of participants, a standing committee, and an administrator at Hamburg.

The Teachers College participant in the study would be the Horace Mann-Lincoln Institute, within which the pilot work of the project has been conducted. It is anticipated that the University of Chicago will continue its present cooperative relationship with this study.

VII. FACILITIES

Data processing facilities appropriate for the study are available at ETS, Columbia University, and the University of Chicago. A 1620 system is available at Teachers College, Columbia University. Large university librarian are available at both Columbia University and the University of Chicago. A large technical library is available at ETS.

VIII. FUNDS

The proposed study is relatively large in scope and time. While it is not possible to present a budget now, the general types of expenditures can be anticipated:

1. Direction and travel
2. Technical assistance
3. Publishing and communications
4. Costs within countries
5. Data processing

We would propose applying to Project English of the U.S. Office of Education for funds. The size of the proposed grant would be in excess of $250,000.

IX. PARTICIPANTS

Several general cultural regions are pertinent to a study of literature, since these regions have developed distinguishable schools of literary criticism within them. Ideally, at least two countries from each of these regions should participate. These regions can be identified by language or by areas, as follows:

1. French-speaking Europe
2. Scandinavia
3. English-speaking countries
4. Italo-Hispanic area
5. Asian countries
6. German-speaking countries

I. E. A.

REPORT OF THE USA COMMITTEE ON THE TESTING OF FRENCH AS A FOREIGN LANGUAGE

Members of the Committee

Professor Wilmarth Starr, Department of Romance Languages, New York University
Mr. F. Andre Paquette, Modern Language Association and American Association of Teachers of French
Professor John B. Carroll, Harvard University (Chairman)
Mr. John L.D. Clark, Harvard University

Introduction

This is the report of a committee convened at the request of Professor S. Bloom of the University of Chicago to outline a provisional proposal for an international investigation of student achievement in French as a foreign language. The proposal represents strictly a USA point of view in the sense that it sets forth possible specifications for the investigation as it might appropriately be conducted in the USA. It is anticipated that the proposal will be reviewed by an international committee and coordinated with proposals received from other participating countries in order to develop a uniform design applicable in all these countries.

The teaching of modern foreign languages in the USA

This section presents a brief summary of the history and current status of the teaching of modern foreign languages in the USA, with particular attention to French, in order to provide necessary background for examining the detailed proposals made as to the design of the study.

The teaching of modern foreign languages did not become a normal and accepted part of the curriculum in the American secondary school until around the middle of the 19th century. At that time, the few secondary schools that existed were attended, for the most part, only by the relatively small segment of the corresponding age groups in the population who intended to go on to some form of higher education. Towards the end of the 19th century and continuing into the 20th, the evolution of the comprehensive high school brought larger and larger sectors of the population into the secondary school. Although total foreign language enrollments increased, the proportion of high-school students enrolled in foreign language courses tended to decrease, since foreign language study was still sought, in the main, by those in college preparatory curricula. The pressure to know a foreign language for any utilitarian purpose is obviously not as strong in the USA as it is in many other countries, e.g., in those countries in which English constitutes a language of wider communication value than the indigenous language. On the other hand, the trend in the USA since World War II has been a gradual increase in the relative numbers of secondary school students electing foreign language study, and for the first time there have been serious and at least partially successful efforts to introduce foreign language study into the elementary school curriculum, even as early as the kindergarten in some cases. These trends are not solely due to the fact that increasing numbers of students hope to attend college; they also appear to stem from genuine interest in foreign language study for its own sake and also for its use in leading to better international communication. The NDEA of 1958 gave an enormous boost to FL study through its provision of government funds for the training of foreign language teachers and for language laboratory equipment in schools. For a general statement on the current status of foreign language study in American life, see W.R. Parker's The National Interest and Foreign Languages (1957).

For the first half of the 20th century, the primary aim of foreign language teaching in the USA was that of producing a reading knowledge of a language. It was believed that such an objective was the only reasonable one to expect of students in the relatively short periods (usually no more than two years) that students customarily studied any one language. There were few teachers who could effectively teach the spoken language. The teaching was chiefly of the "grammar-translation" variety; that is, students were taught rules of grammar and required to apply them in performing written translation from and into the target language and oral translation from the target language. In the public mind, the study of foreign language was regarded as generally unsuccessful and unproductive. In World War II, however, it was dramatically demonstrated that new methods of foreign language teaching, giving priority to mastery of spoken language, could be productive and successful even in periods of two years of study. The revolution in methods of language teaching was so thorough that public attention was brought to the desirability of longer periods of language study in order to attain true and effective mastery of all aspects of a foreign language. Furthermore, methods of language teaching became more refined and effective, especially with the introduction of the language laboratory and other technological aids. Increasing numbers of competent foreign language teachers became available, even at elementary school levels. Nevertheless, progress in foreign language teaching has been uneven. Traditional aims and methods persist and are occasionally stoutly defended in many schools throughout the country, particularly in the less affluent districts. Because the school curriculum is determined largely on a local basis, wide variations in the quality of instruction are found.

Although Spanish is the language which is studied by more secondary school pupils than any other language, French runs a close second, and in some sections of the country or in many school districts, it is elected by more pupils than Spanish or any other language. There is still a tendency to regard French as par excellence the language of culture and of world communication. German is considerably behind French in enrollments, and Russian still trails far behind, even though it is increasing in popularity.

Specifications for testing student achievement in French

Because the study of French can be begun at many different age or grade levels - all the way from the kindergarten to the graduate school - there is a certain arbitrariness about setting up three levels of proficiency to correspond with age levels. Nevertheless, in order to accord with the spirit of the suggested guidelines set forth by the IEA Committee, we recommend that testing be done at three age levels: the 13-14 age level (end of grade 8), the 15-16 age level (end of grade 10), and the 17-18 age level (end of grade 12, pre-university). However, in the selection of the students to be tested, account should be taken of the approximate number of contact hours of instruction they have had. Thus, in testing at the end of grade 8, the sample should be restricted to those who by that time have had a cumulative total of 160-180 contact hours of instruction in French. Further details are shown in the accompanying chart which indicates the degree of competence in each of four basic skills, - listening, speaking, reading, and writing, - at each of the three levels.

No systematic or detailed inventory of either lexicon or syntax has been suggested as the basis for a corpus of testing items, i.e., a list of words and constructions that could be divided into three parts presuming progress from the simple to the more complex. The reason for not doing so is that contemporary language teaching methods in the USA increasingly view language learning as a progressive experience implying enlargement and enrichment of lexicon together with variations upon the basic structure of the spoken language. The irreducible elements, therefore, are the complete sound system which can be covered in a fairly short period of time (and which indeed is automatically subsumed as soon as a reasonable number of words are strung together) and basic syntax which is implicit in connected rational speech of even fairly brief duration. As long as one limits himself to material suitable to interest and age level, he is ipso facto including the sound system and basic vocabulary and syntax. Progress in language learning may be better described, then, by the analogy of concentric rings rather than by travel along a linear distance. In the above analogy the total sound system and reasonable coverage of syntax should be understood as included within the central core of the concentric rings. It is hoped that the illustrations for the identification of levels in the grid will illustrate the above concepts.

On the other hand, the concept of frequency of occurrence of lexical and syntactical elements in normal language production may provide a clue to useful techniques for presenting to the learner items of greater frequency first.

For the matter of lexical inclusion it is suggested that "in two high school years or one college year of instruction it is possible to inculate an "active" (speaking-writing) vocabulary of between 500 and

SKILLS	13-14 age level end of grade 8 160-180 contact hours	15-16 age level end of grade 10 430-450 contact hours	17-18 age level end of grade 12 (pre-university) 700-720 contact hours
LISTENING	Ability to understand an educated native speaker when he is speaking carefully on a subject appropriate to the age level.	Ability to understand conversation of average tempo and sustained presentations including lectures, radio programs, and short films.	Ability to understand with reasonable ease rapid standard speech including classroom discussions, radio programs, recorded plays and full length films.
SPEAKING	Ability to repeat brief meaningful utterances, to read simple passages, aloud, to answer short direct questions, and to describe simple situations; all appropriate to the age level and understandable to a native speaker.	Ability to carry on a conversation with a native speaker without making glaring mistakes, to summarize orally and to answer questions on the content of conversations and presentations which have been heard.	Ability to speak fluently approximating native speech in intonation and pronunciation, in a variety of situations including sustained presentations and classroom discussions of literature.
READING	Ability to understand the meaning of simple prose appropriate to the age level	Ability to read (with glossaries in the FL or with an FL dictionary) prose of average difficulty including selected contemporary plays and short stories in terms of their manifest content.	Ability to read with reasonable ease and with minimal use of lexical aids non-fiction, including magazines and newspapers, and literary materials, including selected classics.
WRITING	Ability to spell and write correctly sentences and short paragraphs appropriate to the age level without glaring errors.	Ability to write directed compositions, resumes, and letters, and to use varied paraphrase techniques with reasonable clarity and correctness as limited by morphology and syntax thus far studied.	Ability to write summaries of oral discussions, compositions on topics of interest with appropriate choice of idiom without glaring mistakes in morphology and syntax.

1,000 words, and a "recognition" (hearing-reading) vocabulary of approximately 1,500 to 2,500. "Language", of course, is more than a body of isolated words that can thus be counted, but these figures give us a basis for significant comparison."[1] These figures also seem to equate with the basic 1,300 words of Le Francais Fondamental.[2] Such a basic vocabulary could be assumed to be covered in about 300 contact hours, or, in the sequence outlined here, at a point about equal to the end of grade 9.

The problem of syntax is an even more arbitrary one. Useful guides are suggested, such as the section "Structures for Four- and Six-Year Sequences" in French for Secondary Schools, Bureau of Secondary Curriculum Development, New York State Education Department, Albany, 1960, or the indexes of the more widely used French "grammars" which generally cover the basic structures of the language in one volume after ordering them in an arbitrary system based upon notions of greater to less frequency and simpler to more complex patterns.

Sample test materials

Sample test items for each of the four skills are given on pages A-58 through A-72. The approximate level of difficulty of each item has been designated, but these are not intended to correspond exactly with the three levels at which the tests are to be given, since the tests to be given at each level should include a range of difficulty.

The items are for the most part of the "objective" variety, and in most cases they can be scored objectively. They have been drawn from a variety of standardized tests now available in the United States.

Additional data to be collected from students

In addition to obtaining student test scores for each of the four language skills, the project should undertake, in so far as possible, to obtain data on background variables which may be thought of as affecting student performance on the tests themselves. It is suggested that a detailed questionnaire be prepared for administration to students at each of the three proficiency levels at about the same time that the tests themselves are being given. The questionnaire would ask the student to indicate the nature of his foreign language experiences, both in the classroom situation, and in such informal contexts as travel in foreign countries, acquaintanceships with foreign visitors, different languages spoken in the home, and so forth.

---

[1] Starr, Wilmarth H., et. al. Modern Foreign Languages and the Academically Talented Student. National Education Association, 1960. p. 79.

[2] Gougenheim, G., Michea, R., Rivenc, P., and Sauvageot, A. L' elaboration du francais elementaire. Paris: Didier, 1956.

Although these informal contacts may be expected to contribute to student achievement in the foreign language to some extent, it is suggested that an analysis of the nature and content of the formal language courses which the student has taken, from the beginning of his foreign language study up to the time at which the tests are administered. will in the majority of cases show a close correspondence with his observed proficiency in the four skill areas. Thus, a major portion of the student questionnaire would be devoted to a description by the student of the language courses which he has taken. The questionnaire would be prepared in such a way as to enable the student to check, from a number of descriptive statements about class organization, methods, and emphasis, the statements which most closely describe the courses which he has had. Examples of items which may be of use in this connection include the following:

1. How many class meetings were there each week?
________one
________two
________three
________four
________five or more

2. How many students were in the class?
________less than 10
________10 to 20
________20 to 30
________more than 30

3. During the class periods, the students would
________usually speak in English unless imitating the teacher in certain drills
________usually speak in the foreign language for drills, conversation and questions

4. For this course your class
a. ________did not use a language laboratory
b. ________used a language laboratory every now and then (not regularly every week)
c. ________used a language laboratory regularly, - that is, a certain amount every week

If you check "c", answer this question too:

How much time each week did your class spend in the language laboratory?
________15 min.
________30 min.
________45 min.
________1 hour
________1½ hours
________2 hours
________2½ hours
________3 hours
________more than 3 hours

5. Did native speakers of the foreign language ever come to talk to the students during class periods?

________no, never
________yes, occasionally
________yes, reguarly

Experience with questionnaire items of this type indicates that quite detailed information about previous language courses may be obtained in an economical manner, both from the standpoint of time required for administration of the questionnaire and the time required for tabulation and analysis of the data obtained.

In addition to the student's formal and informal language contacts, a third variable affecting measured language achievement is student motivation for, and general orientation toward, the study of a foreign language as such. These aspects can be measured with some validity by the inclusion of an appropriate series of items in one part of the student questionnaire. Below are some examples of types of items which have previously been used in this connection:

Directions: For each question, please check the one answer which best describes your feelings:

1. How much do you like foreign language classes compared to your other classes?
   ________I like foreign language classes better than most of my other classes
   ________I like foreign language classes about as much as most of my other classes
   ________I like foreign language classes less than most of my other classes

2. If I had a friend who was a native speaker of the foreign language but who knew how to speak English.
   ________I would usually try to speak with him in the foreign language
   ________I would occasionally try to speak with him in the foreign language
   ________I would really prefer to speak with him in English

Although it is anticipated that some variation in questionnaire format and item phraseology will be required for questionnaires at the three achievement (and age) levels, the essential content of the questionnaires would not vary among the different levels.

Data to be collected from teachers.

A questionnaire should be prepared that could be competed by a sample of teachers at each of the test levels. The questionnaire should consist as far as possible of objective questions in which the respondents would check answers corresponding to their personal opinions or observations, or make ratings along scales showing various degrees of opinion or preference concerning specified issue.

For the purpose of the present document, we think it will suffice to mention some of the more important questions or issues on which information should be obtained:

Objectives of language teaching. What relative importance does the teacher attach to each of the four skills (understanding, speaking, reading, writing)? What relative importance is attached to purely utilitarian aims, such as conversing with speakers of the foreign language, listening to foreign broadcasts, reading books, corresponding with foreign peoples, as contrasted to such objectives as attaining better appreciation of foreign cultures, enjoyment of literature in the foreign language, etc.? To what extent are the objectives of language teaching influenced by the foreign language requirements established by higher schools and college? By external tests and examinations? By pressure from school authorities, parents, etc.? Teachers should be encouraged to give their own opinions on these matters even when they may diverge from "official" points of view as promulgated by school authorities or professional associations.

Methods of language teaching. To what extent, and at what stages, is each of the following aspects of language teaching methodology employed?

Imitation and repetition of words and phrases
Conversational exchanges between teacher and pupil, simple or complex depending on the level of instruction
Use of objects, pictures, actions, gestures to convey meaning
Memorization of dialogues
Practice to perfect pronounciation of FL sounds
"Pattern practice" to perfect mastery of grammatical structures
Use of audio-visual aids, including films, tapes, language laboratory, etc.
Study of grammatical rules and practice by written exercises
Prepared or sight translations of sentences or paragraphs in French into the native language, either orally or in written exercises
Frequent reviews and tests

Again, teachers will be encouraged to give their opinions as to whether the methods they are using are those they would employ if given a completely free hand in the matter.

Conditions of language teaching:

Number and length (minutes) of classroom periods per week; amount of use of the language laboratory; extent of homework assignments
Size of classes; opinions on optimal size of classes
Selection of student for language study. Are all students required to study a foreign language? Is French one of several possible options?
How are students encouraged and guided in their FL study?
Are student grouped for ability? If so, how?
How are texts and other materials selected: by the teacher? by local school authorities? other?

What texts are used? Are they adequate and appropriate?
To what extent does the teacher prepare his own teaching materials?
How many classes, and at what levels, are taught by the teacher?
    Are they all French classes, or are some of them in other languages or in subjects other than foreign languages?
How proficient does the teacher regard himself to be according to MLA standards of proficiency level? What level of teacher training has he attained? Attended an NDEA Institute in French? Had foreign travel or study?

Teacher's evaluation of tests given in this project:

In terms of the objectives and content of the course, are the tests fair?
Are they at an appropriate level of difficulty? Do they present techniques of testing to which the students are unaccustomed?
Do the tests fail to measure important objectives?
What proportion of their students would they estimate would be able to attain, say, 90 percent of the possible points on the tests being administered?

SAMPLE TEST ITEMS

Items for Listening Tests

Listening tests can be prepared at almost any level of difficulty without use of the mother tongue, except for general directions preceding the test. If it is desired to avoid the use of the written language, pictures can be used. Here is an example of an easy pictorial item:

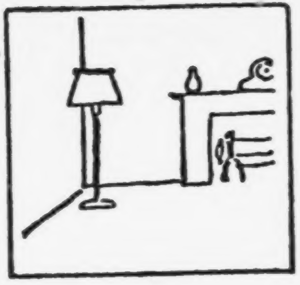
Instructions: You will hear a statement in French. The statement refers to one of the four pictures in your test book. You are to decide which one of the pictures is referred to.

(Spoken on tape:) Il y a des chaises dans le salon.

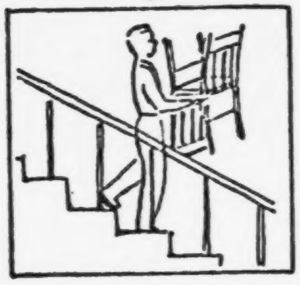
(Pictures in test booklet:)

2

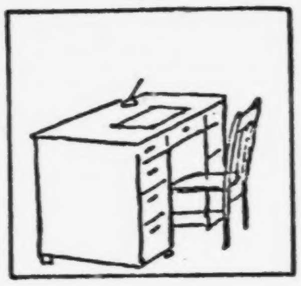
F



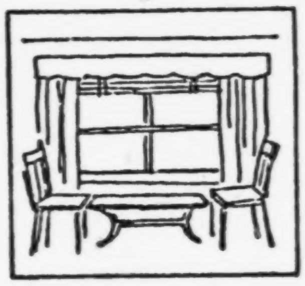
G



H



J



2.......

(From MLA-Cooperative French Tests, Form LA, Listening, Item No. 2, page 4.)

4

If there is no reason for avoiding the use of written language, other types of items are possible. The following are two easy items:

(Instructions both printed in test booklet and spoken on tape):

> You will hear several questions and statements dealing with everyday situations. For each one you are to select the best reply from among the four choices printed in your test book.

(Spoken on tape): C'est après la classe. (Pause)
+ Numéro onze.
Qui est debout à côté du bureau?
Onze. (10 seconds)

(Printed in booklet) C'est après la classe
11
A Sur une chaise
B Des Livres.
C Un drapeau
D Le professeur.

(Spoken on tape): + Numéro onze.
Que faites-vous quand vous quittez l école?
Douze. (10 seconds)

(Printed in booklet):
12
F A trois heures et demie.
G Avez mon ami
H Rien d'important
J Dans l'autobus

(From MLA+ Cooperative Tests, Listening Form LA, Items 11-12, page 5)

Further items of the same general type, somewhat more difficult:

(Spoken script):
Et maintenant, nos annonces publicitaires. (Pause)
+ Numéro trente et un.
Voulez-vous un bon repas? Vous serez bien content chez Dupont. Tout y est délicieuz: potages, viande, légumes et dessert. Le service y est excellent.
Trente et un. (15 seconds)

(Printed in booklet):
Et maintenant, nos annonces publicitaires.
A On parle d'un restaurant.
B On parle d'une pâtisserie.
C On parle d'un petit magasin.
D On parle d'un voyage en avion.

Fairly easy:
Spoken on tape: Nous sommes en voyage. (Pause)
+ Numéro sept.
Vos billets, sil vous plaît.
Sept. (10 seconds)

Printed in booklet: Nous sommes en voyage.

7
A De nationalité américaine
B Les voici, monsieur.

C Non, pas de valises.
D Merci bien, monsieur.

Spoken on tape: + Numéro huit
Mange-t-on vraiment bien au wagon-restaurant?
Huit. (10 seconds)

Printed in booklet:
8 F Non, un wagon-lit.
G Des places vides.
H En traversant deux wagons.
J Le mieux du monde.

From MLA, Cooperative Tests, Form MA, Listening, Items 7-8, page 4.

Fairly difficult
Spoken on tape: Directions: Maintenant nous écoutons deux personnes qui se parlent. Dans chaque cas, choisissez la phrase qui pourrait suivre les remarques de la seconde personne.

(Man A) "J étais justemenet sur le point de tourner à gauche quand ce camion nous a doublés."
(Speed: Moderate-fast. Tone: Shocked.)
(Man B) "Nous l'avons échappé belle! Tu devrais faire attention aux autres voitures, mon vieux."
(Spped: Moderate-fast. Tone: Shocked, then admonishing.)

Printed in booklet:
(a) Je suis moins âgé que toi.
(b) L'accident n'est pourtant pas grave.
(c) Un moment de distraction, c'est tout.
(d) Le camion était arrêté au bord de la route.

From Description of the MLA Proficiency Tests for Teachers and Advanced Students, p. 14.

A difficult "listening" item:
The following dialogue is heard on the tape:
Et maintenant, une petite scène dramatique. (Pause)

Man Non, mon fils Jacques n'a pas assisté au bal hier soir. Il ne songe qu' à travailler et à devenir officier.

Woman Evidemment c'est un garcon très bien.

Man Et sa soeur Annie est aussi très bien, mais enfin, c'est une fille. On ne peut pas autant exiger d'une fille que d'un garcon. Les filles sont faites pour se marier, un point c'est tout. Je n'ai pas voulu qu'elle continue ses études. Elle aurait voulu faire son droit. Le droit! Je lui ai dit: marie-toi et après, si ton mari autorise à faire ton droit, eh bien, ça lui regarde. Et Albert, votre garcon, qu'est-ce qu'il veut faire, lui?

Woman La médecine...Hm! enfin il ne sait pas très bien, au fond. À cet age on est hésitant. La littérature l'attire aussi. Il écrit. Ce n'est pas mal, i...

Man Ecrire! La jeuness d'aujourd'hui ne pense qu'à ca. De mon temp ... enfin, tout ca se résoudra.

Woman Oh! sûrement. Avec la temps, n'est-ce pas..? Et comme le dit mon mari: d'abord passer son baccalauréat, n'est-ce pas?

Man D'abord. Et ne jamais les quitter de l'oeil. Moi, c'est ma méthode. Avec Jacques elle m'a réussi. Respect. Discipline. Culture physique. Il sera officier. Je lui fais apprendre le judo. Il me remerciera un jour. Est-ce c'est votre méthode aussi?

Woman Euh...Oui...en somme...jusqu'à un certain point. Pour le judo en tout cas.

Man C'est la bonne méthode. Mais on ne voit plus nulle part de discipline, de respect, de culture physique. La France meurt, madame, elle meurt, la France.

Woman Peut-être êtes-vous un peu pessimiste...

Man Pessimiste? Il y a cinquante ans qu'elle meurt. Mais naturellement personne ne s'en est encoure aperçu.

Thus is immediately followed by spoken questions:

Voici maintenant les questions (Pause)

Numéro trente-huit.
Selon cet homme, à quoi les jeunes filles devraient-elles se consacrer?
Trente-huit. (10 seconds)

Printed in booklet: Et maintenant, une petite scène dramatique.

38
F A l'enseignement des jeunes.
G Aux bonnes oeuvres.
H A l'emanicipation de la femme.
I Au mariage.

Spoken on tape:

Numéro trente-neuf.
+ Quelle est l'attitude de la femme envers son fils?
Trente-neuf. (12 seconds)

Printed in booklet:

39
A Elle est plutôt tolérante.
B Elle exige trop de lui.
C Elle lui reproche d'être paresseux.
D Elle est fière de ses progrès à l'université.

(Items 38-39, MLA Cooperative Tests, Page 7, Listening, Form MA)

SPEAKING ABILITY TESTS.

Easy Level

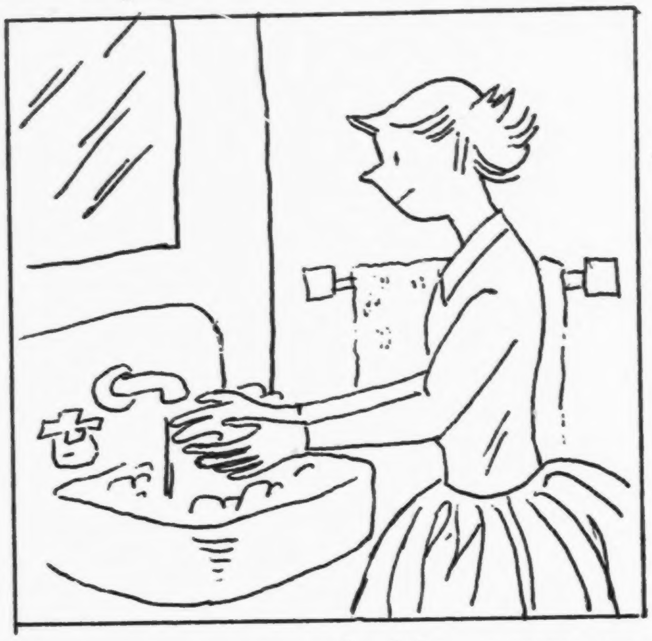
(Directions both printed and spoken:)

You will answer a question about each of several pictures. Answer the question when you hear the signal.

(Spoken on tape:)

Que fait la jeune fille? (1 second, TONE, 8 seconds)

(Printed in booklet):



(The examinee records his answer on tape in the 8 seconds allowed.)

The response is scored on a 3-point scale:

3: Natural and meaningful response, free from error in vocabulary, grammar, and pronunciation (Response may or may not be a complete sentence).

2: Meaningful response with minor error(s) in pronunciation.

1: Response with major error(s).

0: Irrelevant or unintelligible response; not attempted.

Mimicry:

The student is asked to imitate the following sentences spoken on the tape. In scoring, the accuracy of pronunciation of the underlined segments is rated:

Qu'il fait beau (o) Also scored for intonation.
Allons chez eux (ø)
Ils vont à l'école (e)
etc.

Reading Aloud:

The student is asked to read the following passage aloud. In scoring the accuracy of pronunciation of the underlined segments is rated (the underlines do not appear on the text seen by the student):

Deux amis, Paul et Jean, se rencontrent.
--- Ecoute, qu'est-ce que tu fais demain?
--- Demain je vais à onze heures a la gare pour l'arrivée du train.
Je vais rencontrer mon oncle.
--- Tu as un oncle?
--- Oui, c'est un demi-frère de maman. Il est absent depuis six mois.
et je ne le connais pas bien.
--- Qu'est-ce qu'il fait?
--- Il écrit.

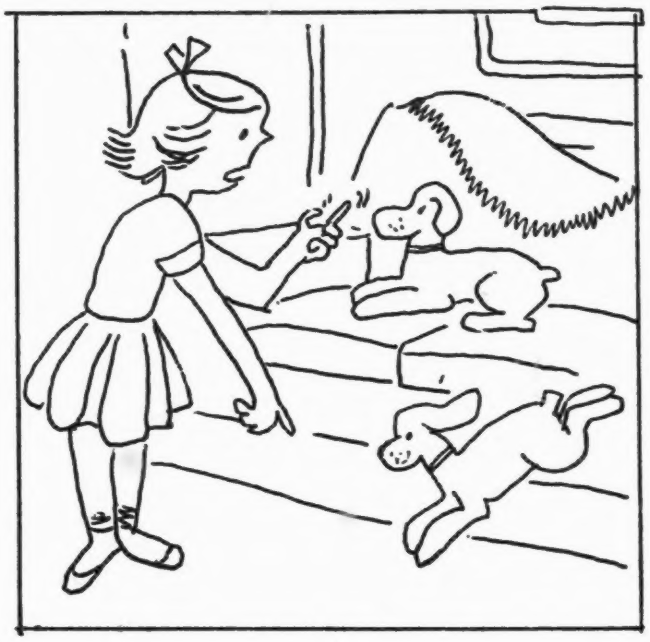
(From MLA Cooperative Tests, French Form L, Speaking)

Difficult speaking ability test:

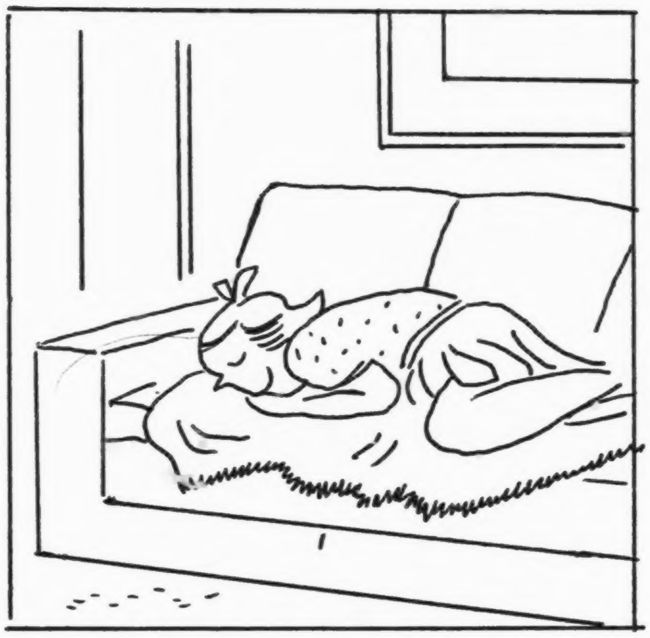
Instructions on tape:

Here you see a simple story told in several scenes. When the signal sounds, tell the story that the pictures tell. Use as much imagination as you like. You may continue talking until you are told to stop.

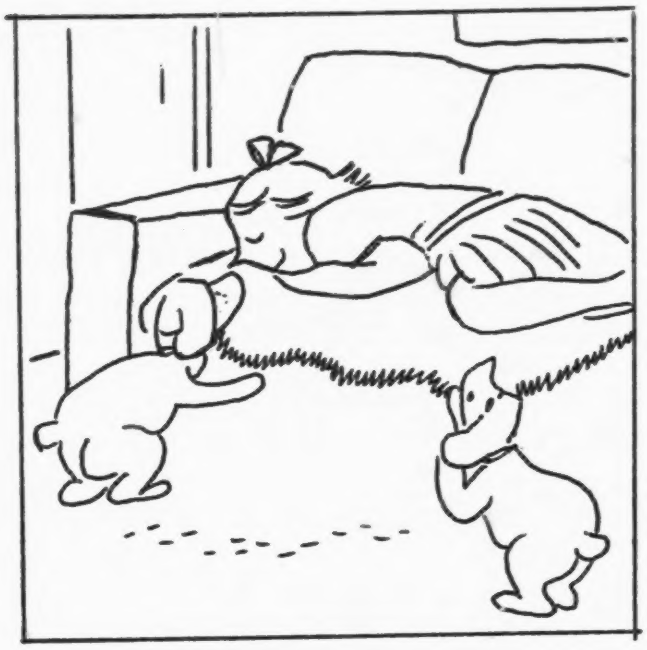
Pictures printed in booklet:



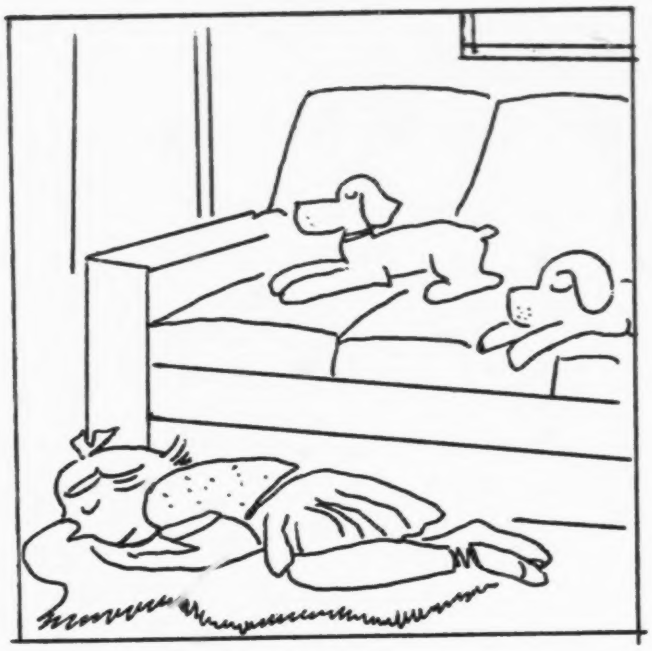
1



2



3



4

Easy: (Items from MLA Cooperative Tests, Form LA, Reading)

Directions: Each of the following sentences contains a blank space ( ) indicating that a word or phrase has been omitted. From the four choices select the one which, when inserted in the ( ), best fits in with the meaning of the sentence as a whole.

1. "Maman, je meurs de faim! Donne-moi donc quelque chose à manger." "Mais voyons, mon enfant, est-ce que tu ( )?"

   A n'as pas ton cahier
   B veux un grand verre d'eau
   C ne peux pas attendre le diner
   D préfères revenir demain

   1 ...

2. Tu cherches toujours tes gants? As-tu pensé à regarder dans ( )?

   F ton tapis
   G ton couvert
   H tes poches
   J tes bas

   2 ...

3. Je vais faire venir le médecin parce que j'ai un ( ).

   A peu froid
   B gros rhume
   C appétit formidable
   D malentendu

   3 ...

23. Ma maison est solide. Elle est bâtie de ( ).

   A glace
   B sable
   C feuilles
   D pierre

   23 ...

24. Chut! Pas de bruit! Parlex ( ) s'il vous plaît.

   F à la fois
   G à haute voix
   H à tort
   J à voix basse

   24 ...

Reading Test Items: Difficulty - Medium to Moderately Difficult

(Items from MLA Cooperative Tests, Reading, Form LA, Items 26-31)

Directions: Each of the selections in this part is followed by several questions or incomplete statements. First read the selection carefully. Then on the basis of the selection decide which choice best answers each question or completes each statement.

"Philippe, va porter cette lettre au bureau de poste," dit la mère de Philippe.
'Oh, maman! Un seul jour de congé par semaine, et je dois faire des courses!"
"Mais qu'as-tu, enfin? Tu seras de retour dans un quart d'heure."

26. Philippe ne veut pas aller au bureau de post aujourd'hui parce qu'il
F ne sait pas où il se trouve
G veut faire plaisir à sa mère
H est facteur.
J préfère s'amuser avec ses camarades

26 ...

27. La mère dit que Philippe pourra aller au bureau de poste et revenir en
A un seul jour
B trente minutes
C quinze minutes
D cinq minutes

27 ...

28. Aux Etats-Unis, cette conversation aurait lieu probablement un
F lundi
G mardi
H vendredi
J samedi

28 ...

Le matin, ma mère me réveille de très bonne heure. Je m'habille rapidement, et je sors avec mon petit sac sur le dos, et le bout du nez et les oreilles bien cachés dans une grosse écharpe de laine, car on gèle chez nous à cette époque de l'année. J'arrive presque toujours avant les autres. J'entre dans la salle encore vide, et je m'assieds à mon pupitre.

29. Cette scene se passe au temps de l'année ou il fait
A doux
B chaud
C froid
D bon

29 ...

30. Cet enfant se rend
F au marché
G au bureau
H à l'eglise
J à l'école

30 ...

31. L'enfant arrive généralement en
A dehors B arrière C retard D avance

Reading Test Items: Fairly Difficult
(Items from MLA Cooperative Tests, Reading, Form MA, Page 17, Items 33-35)

Depuis leur mariage, voilà sept ans, Lucia et Miguel sont aux prises avec le même probleme: régulièrement, les journaux annoncent qu'il a décidé de ne plus combattre, et qu'elle ne fera plus de cinéma. Puis, aussi régulièrement, on écrit que Luis Miguel reprend la cape, pour une saison, et, aussitôt, que Lucia fera peut-être son nouveau film. Mais ce pauvre chantage de la femme angoissée ne réussit pas. Après quelques pourparlers avec un metteur en scène, Lucia comprend vite que son mari descendra dans l'arène de toute facon, et que même la perspective de voir sa femme se replonger dans l'univers du cinéma ne peut l'arrêter.

Quelle est la profession de Luis?
- A Il est journaliste
- B Il est étoile de cinéma.
- C Il est toréador.
- D Il est metteur en scène.

Que pense Lucia du métier de Luis?
- F Elle en est enchantée.
- G Elle le trouve inquiétant.
- H Elle aime toute la publicite que lui donnent les journaux.
- J Elle le préfère au cinéma.

Que savons-nous de Lucia et de Luis?
- A Tous deux fréquentent le cinéma assidument.
- B Ce sont des chanteurs de grand talent.
- C Ce sont un mari et une femme que se battent mutuellement.
- D En dépit de leur amour l'un pour l'autre, Luis reprendra son métier.

"Eh bien, ton ami Paul a de nouveau refusé de faire partie de notre équipe de hockey."
"Ne t'en fais pas! Il finira par accepter. C'est qu'il aime à se ( ).
- F tirer d'affaire
- G montrer tout de suite obligeant
- H fier aux autres
- J faire prier un peu

Reading Test Items: Difficult
(From MLA Cooperative Tests, Form MA, Reading, Page 19, Items 48-50)

Il baissa la glace de la portière. La fraîcheur nue de cette époque de l'année, l'odeur de la terre humide et celle, rassurante, des bêtes de fer me, une cloche grave entendue à travers des bois dont les feuilles jaunissaient.... C'était comme une eau froide et noire qui montait: Marc en fut submergé. Le soir tombait sans réverbères, sans fenêtres allumées: un soir sans hommes, mais peuplé, bruissant, souverain, où March ne se sentait qu'un étranger fragile.

De quelle saison s'agit-il?

F Le printemps
G L'hiver
1 H L'automne
J L'été

Ou se passe cette scene?

A Près d'un village
B Sur un terrain de jeux
C Dans une rue de petite ville
D Au bord de la mer

L'endroit lui

F rappelle son enfance
G paraît bien agréable
H inspire confiance
J est peu sympathique

Tests of Writing Ability

Easy:
Directions: Rewrite each of the following sentences so that the word or words printed on the line below it fit correctly into the new sentence. Make all changes necessary to produce a correct French sentence, but do not add any words and do not leave any out. Do not change the tense of the verbs.

39. Cette nouvelle voiture est rapide et assez dangereuse.
______________________________avion est______________________________
____

40. Tu crois que tu peux y aller tout seul? Tu ne connais pas le chemin!
Vous____y aller tout seul?
____le chemin!

(From MLA Cooperative Tests, French, Form LA, Writing, Page 22, Items 39, 40)

A more difficult, "directed writing" test; it will be noted that the directions for this test require the use of the native language.

Directions (58): On the separate answer sheet, write in French in connected paragraph form an account for your friend of what you have been doing since he left your house this morning and what your plans are for the rest of the day.

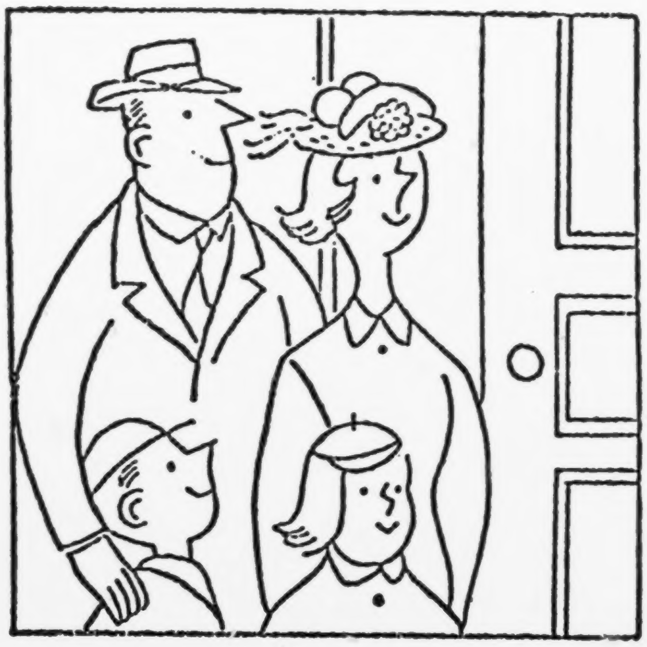
Tell your friend:

a at what time you got up in the morning
b what you had for breakfast
c that you stayed home until noon
d that you spent the whole morning studying mathematics
e that you went to school by bus at 12:15
f that you found the examination easy but too long
g that after returning home you will play tennis with some friends
h that after dinner you intend to listen to French records
i that you have to review everything that is difficult for the French examination
j that by so doing you hope to pass with good marks

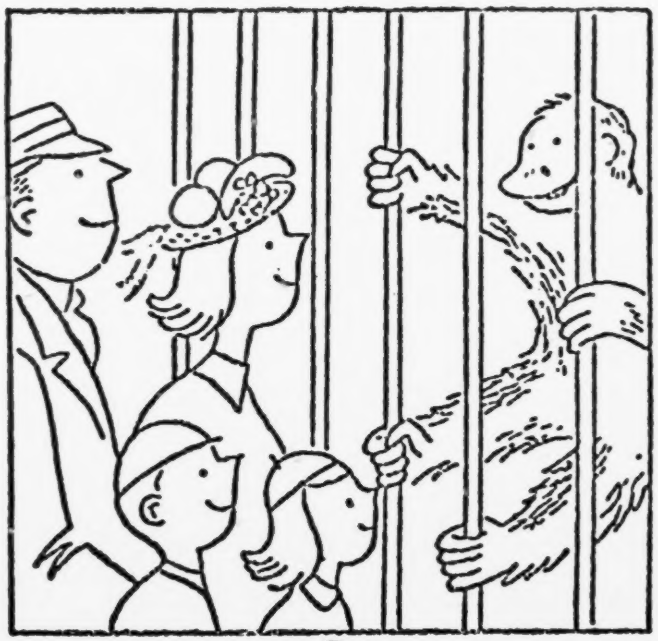
(From the University of the State of New York, Regents High School Examination, French III)

Writing Test: Fairly difficult

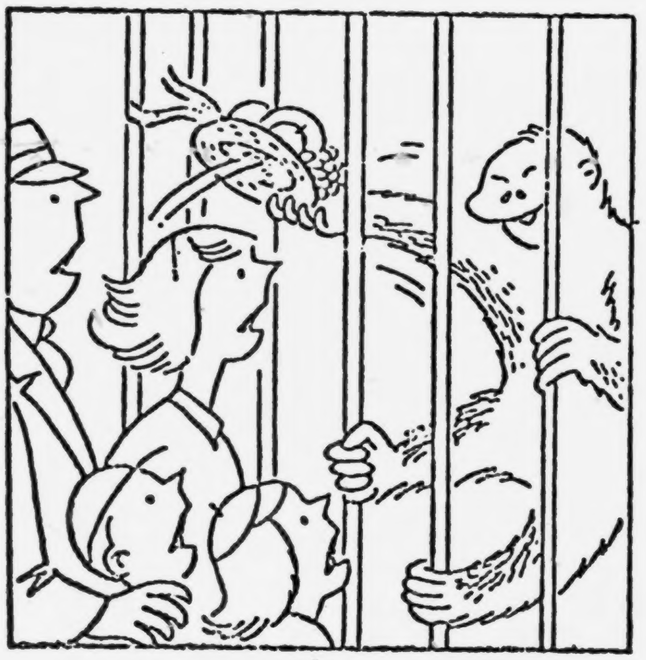
Directions: Write a paragraph telling the story suggested by the four pictures below:



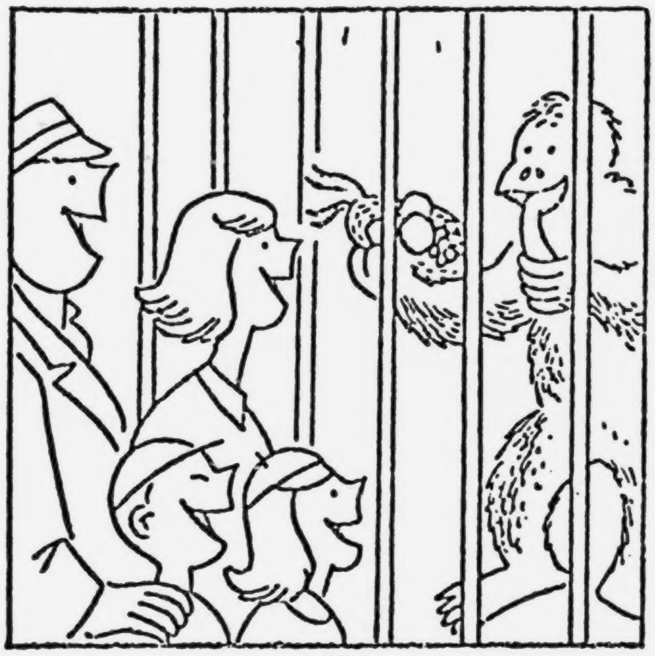
1



2



3



4

APPENDIX B

Examples of Newsletters

IEA/B/55
June 30th, 1968

IEA NEWSLETTER

Dear Colleague,

1. Pre-testing

(a) Cognitive measures

During the past few months you have been inundated with a great amount of pre-testing requests. Unfortunately, some of the tests you received only in May. Thus, in those countries where the school year ends in April or May, pre-testing will have to be delayed until August or September. Appendix I of this Newsletter gives the dates by which item analyses have been or will be returned by various countries to Hamburg. The last row on the chart gives the very final dates by which item analyses in the different subject areas should now be returned to Hamburg. The International Committees will then meet in order to produce the penultimate forms of the tests. These will then be sent out for comment to National Centers and modified in the light of comments received. The final tests therefore should be ready some three months after the date of the return of item analyses as stated at the bottom of the chart. Thus it may be February or March before some of the tests are ready. This will mean that we shall have to change the original schedule for the dry run.

(b) Affective and descriptive measures

The form of pre-testing the affective and descriptive measures has been slightly different from that of the cognitive measures, and is being carried out in two or more rounds. The following is a progress report on the various measures:

(i) Science

You have already received the report of the preliminary trial of Science attitude scales (IEA/ATT/SCI/3). The testing is now in its second round, and the scales "The Place of Science in the School and in the World" and "Interest in Science" are in their second round of testing in all countries at present. The descriptive scale "Science in our School" is now in its first round of testing in all countries.

(ii) Literature

There are two scales, (a) transfer and (b) interest, and the first round results have been analyzed and the second round will go out to all countries within two or three weeks.

(iii)French

The first round concerning ethnocentrism, attitudes the French people and preference for won country have been analyzed and reported in IEA/ATT/FR/5. The second round is now out in all countries and concerns attitudes towards the French culture.

(iv) English

The results of the first round of piloting proved to be unsatisfactory. The problem is being rethought, and we are hoping to produce a general "Interest in English" scale. This will be circulated to all countries participating in English during the next few weeks.

(v) Civic Education

The following instruments have been produced: C.Q.1, C.Q.2, C.Q.3, C.Q.4, D.V.1, D.V.2, C.V.1, C.V.2, C.V.3, C.V.4, C.V.5, and C.V.6. It was hoped that these instruments would be piloted in the United States, England and Germany. So far, some results only have been received from England and Germany. It is expected that from the results of the pre-piloting a new set of instruments can be produced for full pre-testing by all countries participating in Civic Education in September of this year.

(vi) General

So far, four attitude scales, (a) I like school, (b) self esteem, (c) achievement motivation and (d) stereotyped thinking, and one descriptive scale - school environment (containing two dimensions, authoritarian/permissive and subject/pupil orientation) have undergone a first pilot round in England. Revised forms are now being produced and will be circulated to all countries for pre-testing within the next few weeks.

It is hoped that the final pre-testing in all countries of all attitudinal and descriptive measures will take place in September, and that final instruments will be ready by December.

2. Delay in pre-testing

Because we were late in getting tests out to National Centers, there has been a delay in getting results back to Hamburg. According to the attached list (Appendix I), it is clear that only Science, Reading Comprehension and Literature will be able to follow a fairly normal procedure and be ready by December. It is equally clear that delays in the pre-testing of French, English and particularly, Civic Education are occurring, which means that the original dates for the International Committee meetings cannot be adhered to. Most of the pre-testing for French and English will have to take place in the period September to November, which means that the meetings will not be held until later in the year.

The English Committee must still produce the pre-tests for speaking and writing, and it is hoped that these will be ready in time to be administered sometime towards the end of 1968. In the case of Civic Education, it is possible that even longer delays may occur although it is hoped that pre-testing of both cognitive and affective measures can be carried out before the end of 1968. Similarly, many of the attitude and descriptive scales (both for specific subjects and across the board), as well as some questionnaires, will have to have a final pre-testing in September/October, and, hopefully, they will then be ready by the end of the year. It is clear that these delays have implications for the future program.

New deadlines

All of the affective and descriptive measures will continue to be pre-tested during the year. In recent weeks various instruments have been sent to you for pre-testing and in the covering letter you have been requested to indicate the date by which you can return the punched cards to Dr. Choppin in London. It is recognized that in many countries it is too late to test during June/July because of examinations, end of term activities, etc. Thus the instruments will have to be pre-tested in August/September/October. I would appreciate it if Heads of National Centers would respond to Hamburg indicating when punched cards may be expected to arrive on Dr. Choppin's desk.

As far as the cognitive tests are concerned, the situation is as follows:

(i) Science

The International Science Committee Meeting is taking place in Hamburg from July 1st to 10th, and it is hoped that penultimate drafts of the final tests will be produced. It may be necessary to hold a small sub-committee meeting later concerning the Supplementary Science test materials (again it is anticipated that certain National Centers will be pre-testing these Supplementary Science Tests in September. Could this please be confirmed to Hamburg.) Every effort should be made to have the item analyses for the Supplementary Science Tests back in Hamburg by September 30th.

(ii) Word Knowledge, Reading Speed and Reading Comprehension

According to the information sent in by National Centers, it is anticipated that the item analyses for all of these pre-tests will have arrived in Hamburg by July 15th, 1968. They will then be collated and sent to Professor Thorndike for August 1st.

(iii) Literature

For the multiple choice booklets (1 to 6) it is now anticipated that all item analyses will be returned to Hamburg by July 31st 1968.

(iv) French as a foreign language

It is now anticipated that all item analyses can be returend to Hamburg by October 31st, 1968.

(v) English as a foreign language

It is now anticipated that all item analyses can be returned to Hamburg by October 31st, 1968.

(vi) Civic Education

It is now anticipated that all item analyses can be returned to Hamburg by October 31st, 1968, or possibly slightly later.

(vii) Questionnaires

It is anticipated that comments* and frequency distributions, etc. can be returned to Hamburg by October 1st, 1968.

(viii) National Case Study Questionnaire

Unfortunately we are experiencing considerable delay in the production of the National Case Study questionnaire. It is, however, hoped that this can be sent to you within the next six weeks. All that will be required on this questionnaire will be comments.

* It would be appreciated if comments could be back by the end of August.

3. Amendments to IEA/B/53

IEA/B/53 - a chart indicating which pre-tests are being administered in which countries - was circulated on May 21st, 1968. Since then we have been notified of the following amendments:

Australia - not testing Word Knowledge Population I
- not testing Reading Comprehension
Chile - not testing Civic Education
India - testing Reading Speed Populations I and II
- not testing Science Populations IV and IVS
- not testing Supplementary Science
Iran - not testing Literature open ended
Scotland - testing of SST 2 and 3 and French Population I doubtful
Thailand - not testing Supplementary Science

4. Future program for dry run and full testing

It would appear that two alternatives are possible:

(a) Despite the delays in pre-testing in English, French and Civic Education, an attempt is made to expedite the final stage of test construction and produce tests in these subjects in time for the dry run in 1969. In the case of the other subjects, there are no problems in maintaining the original schedule.

(b) To ensure that the best possible instruments are prepared, more time is devoted to finalising the test construction, that is, full time for comments and modification, for French, English and Civic Education. This in turn implies that they would not be ready for the dry run in 1969, but could be available for such use in 1970.

It is suggested that if alternative (b) is chosen, then the future program be amended to incorporate a further stage. Stage 2 would consist of the dry run in 1969 for Science and Mother Tongue, with the full testing in 1970. Stage 3 would consist of the dry run and full testing for French, English and Civic Education at later dates to be agreed bu the Council.

5. Reasons for the dry run

I have received several queries about the purposes of the dry run. There are in fact two main purposes, which are:

(i) to test procedures
and (ii) to test computer programs

Re. (i): It should be possible within two to three months of receiving the data for a population from a country to be able to raise queries with the country about possible errors in procedures (for example, incorrect coding, punching erros, etc.)

Re. (ii): The dry run will enable all programs to be tested with real data. This, it is hoped, will enable as many snags as possible to be removed.

6. Literature cross-over study

In the spring of this year a study was made in England and the United States on multiple choice versus open ended items. Multiple choice items proved to be as reliable as open ended items and also appeared to be measuring the same thing. Although the reliabilities for the prose texts were much higher then for the poetry texts, they are still not very high (that is, about .65 for the prose and .22 for the poetry). A full report (IEA/TR/6) will be circulated within the next few weeks.

7. Finance

I am pleased to be able to inform you that the Volkswagen Foundation has granted us $200,000 for Stage 2 of our work.

8. Proposed analyses for full testing

Considerable work has been undertaken in the last few months in the planning of the proposed analyses. Two documents are at present being produced, one in lay terms and the other in technical terms, describing in some detail the proposed analyses. These will be circulated to all IEA members in September.

9. Proposed administrative arrangements for full testing

On May 27th you were sent a copy of IEA/B/54 which gives an outline for the proposed administrative arrangements. It is most important that we get the reactions from National Centers to this document as soon as possible. May I, in your own interest, urge you to send in your comments to Dr. Postlethwaite in Hamburg as soon as possible.

10. Development of IEA data processing plans

Dr. Bruce Choppin has begun work in London on the development of certain computer programs. Initially these are designed to handle the present pre-testing work and the development of the attitudinal and descriptive measures, but it is expected that towards the end of this year this work will be expanded to include more general statistical programs for use in the next stage of IEA. It is difficult at the moment to offer these programs for general use by National Centers, because there are no supporting documents.

11. Amendments to IEA Statutes

In order that IEA becomes exempt from corporation tax in Germany, it has been found necessary to amend the IEA Statutes. The amendments are as follows:

(a) Article 2 is now divided into two paragraphs.
Insert "(a)" before the first paragraph.
Insert "(b)" The co-ordination center of the Association is situated in Hamburg (Federal Republic of Germany). The co-ordinator is responsible for the day-to-day management of the Association."

(The explanation for this is that in order for IEA to be exempted from corporation tax in Germany, it must be regarded as a socially useful organization ("gemeinnutzig"), and to be so recognized must either have its official seat or its management seat in Germany.)

(b) Insert in the last paragraph of Article 4 after " . . . forfeit all rights in the Association.":

"They receive merely the amount of their paid-in capital and the fair market value of their contributions in kind, if any."

(c) Insert a new Article after Article 22:

"22a. In the event of profits being made by the Association, they may be used only for the aims stated in Article 1 of these Statutes. The members may take no share in the profits, and in their capacity as members may receive no gift, donation, grant or loan from the Association. The Association may not incur administration expenses foreign to the aims of the Association or pay disproportionate amounts (for example, salaries, allowances, honoraria, etc.) to any one person."

(d) Delete "its assets" at the end of the second paragraph of Article 24 and insert the following:

" . . . the assets remaining after the repayment of (a) the paid-in capital, if any, and (b) the fair market value of the members' contributions in kind, if any. The remaining assets may only be used for aims in accordance with Article 1 of these Statutes.

> The decision of the Council regarding the use of the remaining assets requires the approval of the competent tax authority."

A formal decision will be required at the December Council Meeting to amend the Statutes. (Will Council members please regard this as a formal notification (as required by Article 24 of the Statutes) of amendments to the Statutes. It is still hoped that we may get a special exemption from the German tax authorities without changes to the Statutes, but should this fail we must be in a position to change the Statutes in December.)

12. IEA Publications

You will be pleased to hear that a short version of the IEA Mathematics volumes has appeared in Germany under the title: "Leistungsmessung in der Schule: Eine Internationale Untersuchung am Beispiel des Mathematikunterrichts". It is published by the Verlag Moritz Diesterweg, Frankfurt and costs DM 14,--. A short version in Swedish, written by Mats Hultin, will be coming out in the very near future.

Yours Sincerely,

Torsten Husén

ITEM ANALYSES RETURN

Country	Science	Science Supplementary	Word Knowledge	R.C.and Rdg.Speed	French as a Foreign	English Language	Literature Multi,	Literature Open	Civic Education
Australia	15.6	20.6	22.6	-	-	-	-	-	-
Belgium	15.6	-	-	21.6	-	-	15.7	-	-
Chile	30.6?	1.8	30.6?	30.6?	(Por.IV)15.9	(Pop.IV)15.9	30.7	-	Not doing
England	15.6	17.6	Sent to Thorndike	10.6	24.6	-	17.6	24.6	29.7
Red.Rep.of Germany	30.6	30.6	-	-	-	30.7	-	-	30.7
Finland	15.6	20.6	29.6	29.6	-	27.7	15.6	29.6	30.7
France	Strike	-	Strike	-	Strike	-	-	-	-
India	15.6	-	21.6	21.6	-	-	-	-	-
Iran	15.6	30.9	1.6	22.6	30.9	1.10	15.6	1.10	1.10
Italy	May 69	May 69	May 69	15.10	May 69	May 69	May 69	May 69	May 69
Japan	1.6	25.6	?	-	-	-	-	-	-
Netherlands	25.6	-	15.7	15.7	15.10	15.10	-	-	-
Poland	1.6	31.10	15.6	15.6	-	-	-	-	-
Scotland	25.6	SST 25.6	22.6	21.6	?	-	-	-	-
Sweden	15.6	15.9	15.9	15.6	R&L W&S 15.6 15.9	20.9	15.9	15.9	15.9
Thailand	22.6	Not doing	22.6	22.6	-	27.7	-	-	Not doing
U.S.A.	18.6	21.6	1.6	1.6	25.6	-	28.6	-	28.6
Final Return to Hamburg	25.6	30.9	15.7	15.7	31.10	31.10	31.7	-	31.10 or possibly later

APPENDIX B-2

IEA/B/74

31st December 1968.

IEA NEWSLETTER

Dear Colleague,

May I take this opportunity of wishing you all the best for 1969 - our dry-run year. A great deal has happened in the last three months and without going into great detail, I shall try to summarize the points of interest to all persons involved in IEA.

1. New Centers

At the IEA Council meeting in Hamburg in December 1968, two new research institutes were admitted to the IEA Council. These were:

a) The National Institute of Pedagogy, Budapest, Hungary.
b) The Curriculum Unit of the Department of Education, Wellington, New Zealand.

This now makes 19 research institutes actually participating in IEA and the IEA Council decided that for the time being the number of research institutes should be restricted to twenty, but that if further applications are received, the Council should review its policy of admission of further centers in terms of the general enlargement and costing of IEA's work.

2. Change of membership of Standing Committee and Bureau

The terms of office of Professor Bloom (Chicago), Dr. Walker (Edinburgh) and Professor De Landsheere (Liège) on the Standing Committee have come to an end, and Professor Bloom indicated his desire to resign from the Standing Committee and also from the Bureau. Dr. Walker indicated his desire to resign from the Standing Committee. The Council, therefore, elected Professor Foshay, (New York) and Dr. Ayman (Tehran) each for a two-year period, and re-elected Professor De Landsheere to the Standing Committee for a further two-year period. The Standing Committee elected Professor De Landsheere to the Bureau for a one-year term of office. These changes of membership formally take place on April 1st, 1969.

3. New Chairman of the International Literature Committee

Professor Foshay has indicated his desire to resign as Chairman of the International Literature Committee, although he will continue as a member of the Committee. The Bureau has invited Professor Alan Purves of the University of Illinois to be the new Chairman, and Professor Purves has accepted. His appointment takes effect from January 1st 1969.

4. Revision of IEA Statutes

Certain minor changes have been made to the IEA Statutes (see June

1968 Newsletter) by the Council so that IEA can be recognized as corporation tax free in the Federal Republic of Germany.

5. Stage 2 Tests

The test instruments for Stage 2 have now been completed. They are as follows:

No.	Test	No. of Items	Time (Mins.)	Optional	Bulletin of Erratum Sheet
Science					Pink Bulletin and erratum sheets 1 and 2 (pink)
IEA/1A	Pop.I	20	30		
IEA/1B	Pop.I	20	30		
IEA/4A	Pop.II	40	60		
IEA/4B	Pop.II	40	60		
IEA/10A	Pop.IV	36	60		
IEA/11B (P-U)	Pop.IV	30	60		
IEA/15	IVS Bio	40	60	Optional	
IEA/16	IVS Che	40	60	Optional	
IEA/17	IVS Phy	40	60	Optional	
IEA/19	II Practical	9	90	Optional	
IEA/20	IV Practical	7	90	Optional	
IEA/21	IV " (Bio)	5	90	Optional	
IEA/22	IV " (Che)	3	90	Optional	
IEA/23	IV " (Phy)	7	90	Optional	
Reading Comprehension					Blue Bulletin and erratum sheet no.
IEA/3C	Pop.I	21	25		
IEA/3D	Pop.I	24	25		
IEA/3P	Pop.I(Practice)	40	10		
	Reading Speed				
IEA/3J	" "	40	10		
IEA/7C	Pop.II	26	50		
IEA/7D	Pop.II	26	50		
IEA/7P	Pop.II(Practice)	40	10		
	Reading Speed				
IEA/7J	" "	40	10		
IEA/14C	Pop.IV	25	50		
IEA/14D	Pop.IV	29	50		
Word Knowledge:					
IEA/2F	Pop.I	40	10		Yellow Bulletin
IEA/6F	Pop.II	40	10		" "
IEA/13F	Pop.IV	40	10		" "
Literature					Blue Bulletin and erratum sheets no. 1 and 2 (blue)
IEA/8X		39	50		
IEA/8Y		38	50		
IEA/8Z		39	50		
IEA/9W		37	50		
IEA/18		1	60	Optional	

No.	Test	No. of Items	Time (Mins.)	Optional	Bulletin or Erratum Sheet
Questionnaires			See Manual 1 for times of these		Yellow Bulletin and erratum sheets no. 1, 2, 3, and 4 (yellow)
Student General					
IEA/2G	Pop.I	40			
IEA/6G	Pop.II	41-79			
IEA/13G	Pop.IV	41-79			
Student Science:					
IEA/5S	Pop.II	28			
IEA/12S	Pop.IV	28			
Student Literature			See Manual 1		Yellow Bulletin and erratum sheets no. 1, 2, 3 and 4 (yellow)
IEA/9Q		35			
IEA/TQ1	Teacher General	38			
IEA/TQ2	" Science	21			
IEA/TQ3	" Reading Comprehen.	23			
IEA/TQ4	" Literature	38			
IEA/SQ1	School Questionnaire	33			
Attitude and Descriptive Scales					
General:					
IEA/2H	Pop.I	36			
IEA/6H	Pop.II	36			
IEA/13H	Pop.IV	36			
Science:					
IEA/5T	TOUS Pop.II	15			
IEA/12T	TOUS Pop.IV	15			
IEA/1K	Pop.I	22			
IEA/5K	Pop.II	48			
IEA/12K	Pop.IV	48			
Literature:					
IEA/9R		20			Erratum Sheet No. 2 (blue)

The optional tests will be given by the following National Centers:

Science IVS: England, Fed. Rep. of Germany, Hungary, Iran, Japan, New Zealand, Scotland, Sweden.
Science Practical:
Pops. II and IV: Chile, England, Iran, Italy, Japan, New Zealand, Scotland.
Pop. IVS: Chile, England, Iran, Japan, New Zealand, Scotland.
Literature IVS: National Centers have not yet indicated whether or not they will be participating in this. Could Heads of Centers please write to Dr. Postlethwaite as soon as possible?

State of Attitude and Descriptive Scales

As a result of a long discussion at the Council meeting, it was decided that the work on the F scale and the self-esteem scale should be dropped. Some changes to the Science scales (in the yellow bulletin) have been made and appear in Erratum Sheet No. 1 (yellow).

At the beginning of January the following scales have been sent out:

	No. of Items	Erratum Sheet
IEA/2H, 6H and 13H General for Pops. I, II and IV	36 each	No. 3 (yellow)
IEA/9R Literature	20	No. 2 (blue)

Dr. Choppin has also prepared accompanying notes for each set of scales indicating the intent of each item. National Centers should take care to refer to these notes when translating the items. (This also applies to National Centers in countries where the mother tongue is some form of the English language.)

6. Dry Run for Stage 2

The instructions for the administration within countries are included in Manuals 1, 2 and 3 in the Green Bulletin produced in December 1968.

The following table indicates at which time National Centers are carrying out the Dry Run Testing:

National Center	Population I		Population II		Population III		Pop'n IV & IVS	
	Dry Run	Main Testing	Dry Run	Main Testing	Dry Run	Main Testing	Dry Run	Main Testing
Australia	-	-	7/69	7/70	-	-	7/69	7/70
Belgium	3/69	3/70	3/69	3/70	-	-	3/69	3/70
Chile	10/69	10/70	10/69	10/70	-	-	10/69	10/70
England	6/69	6/70	3/69	3/70	3/69	3/70	3/69	3/70
Fed. Rep. of Germany	5-6/69	5-6/70	5-6/69	5-6/70	-	-	5-6/69	5-6/70
Finland	3/69	3-4/70	3/69	3-4/70	3/69	3-4/70	11/69	1-2/70
France	-	-	-	-	-	-	4/69	4/70
Hungary	5/69	5/70	5/69	5/70	5/69	5/70	5/69	5/70
India	9/69	2/70	9/69	2/70	-	-	5/69	2/70
Iran	2/69	2/70	4/69	4/70	4/69	4/70	2/69	2/70
Italy	2/69	3-4/70	2/69	3-4/70	-	-	2/69	3-4/70
Japan	4/69	4/70	4/69	4/70	-	-	10/69	10/70
Netherlands	Dry run: Planned for March, if not possible Oct. 1969 Main testing: March 1970							
New Zealand	-	-	7/69	7/70	-	-	7/69	7/70
Poland	5/69	5/70	5/69	5/70	-	-	3/69	3/70
Scotland	6/69	5-6/70	6/69	5-6/70	-	-	6/69	3/70
Sweden	5/69	5/70	5/69	5/70	5/69	5/70	5/69	5/70
Thailand	6/69	12/69	6/69	12/69	7/69	12/69	7/69	12/69
U.S.A.	4/69	4/70	4/69	4/70	-	-	4/69	4/70

The next table gives the final commitments made by National Centers to Stage 2 testing:

	Science					Reading Comprehension				Literature			
	I	II	III	IV	IVS	I	II	III	IV	II	III	IV	IVS
Australia	x	x		x									
Belgium		x		x		x	x		x	x		x	
Chile	x	x		x	x	x	x		x	x		x	
England	x	x	x	x	x	x	x	x	x	x	x	x	x
Fed. Rep. of Germany	x	x		x									
Finland	x	x	x	x		x	x	x	x	x		x	
France				x									
Hungary	x	x	x	x		x	x	x	x				
India	x	x		x		x	x		x				
Iran	x	x	x	x	x	x	x	x	x	x	x	x	x
Italy	x	x		x		x	x		x	x		x	x
Japan	x	x		x	x								
Netherlands	x	x		x		x	x		x				
New Zealand		x		x	x		x		x	x		x	x
Poland	x	x		?x		x	x		?x				
Scotland	x	x		x	x	x	x		x				
Sweden	x	x	x	x	x	x	x	x	x	x	x	x	
Thailand	x	x	x	x		x	x	x	x				
U.S.A.	x	x		x		x	x		x	x		x	x

7. Stage 3 Tests

Work is continuing with the preparation of instruments for French as a foreign language, English as a foreign language and Civic Education. Penultimate versions of reading comprehension and listening comprehension French tests for Populations I, II, IV and IVS have now been prepared and these will be repretested between February and May 1969 in England, Iran, Italy, the Netherlands Sweden and the U.S.A. New writing and speaking tests will be prepared and will be prepared and will be pre-tested at the same time in the same countries, as well as in Chile. The final versions of all French instruments will be prepared in early July at the next meeting of the International French Committee.

The item analyses for Civic Education are nearly all in, and proposed final tests will be prepared at the end of February by a sub-committee of the International Civic Education Committee. More work will be required in pretesting the attitude scales in Civic Education throughout 1969. In the meantime, the Civic Education Committee has produced a revised rationale for its testing in "The Assessment of Civic Education" (IEA/CIV/12).

The item analyses for English as a foreign language are coming in slowly. However, the writing and speaking tests have not yet gone out to National Centers, but it is hoped that they will go out in the course of the next three weeks. Fortunately, since the dry run testing for this subject is not to take place until 1970, we have slightly more time to complete the pretesting.

8. Mathematics Data Bank

The data in the mathematics data bank can now be processed on the IBM 360 systems as well as on the 7094 Computer at Chicago. The data will be made available by IEA both at the Stockholm computer and the Columbia University computer. All requests should be sent to the Chairman of IEA:

Professor T. Husén,
I.E.A.,
c/o Unesco Institute for Education,
2000 Hamburg 13
Feldbrunnenstrasse 70.
Federal Republic of Germany.

All approved requests will be carried out by the IEA Data Processing personnel at cost.

9. IEA Item Bank

IEA has, in general, pretested about 3-4 times as many items as go into its final tests with the result that many 'good' items are not used. It is considering the possibility of establishing an item bank where these would be available for use by bona fide researchers.

10. IEA Personnel and Locations

John Hall of the N.F.E.R. joined the IEA staff on January 1, 1969, and will be working with Bruce Choppin on the data processing. They will continue the development work until the end of March 1969, in London and move to New York at the beginning of April. Mr. Richard Russell of London will be working for the first six months of 1969 as a programmer.

Neville Postlethwaite has been given the title of "Executive Director".

It is still unclear whether the IEA Headquarters will remain in Hamburg. The Unesco Institute will have the whole matter of IEA's accommodation discussed at its Governing Board meeting in April, whereafter a decision will be made.

11. IEA Visiting Scholar Center

The Swedish Government has very kindly given IEA a suite of offices in the Wenner Gren Center in Stockholm. These offices will be used partly by the Swedish National Center and partly by IEA International. At the same time educational research scholars working on an IEA problem will be invited to Stockholm for short periods of time (1 - 6 months) as visiting scholars, and will also be provided with personnel accomodation in the Wenner Gren apartment complex.

As usual, if you have any queries on the content of this newsletter, or if you should require further information, I would appreciate it if you would contact either Dr. Postlethwaite or me.

Yours sincerely,

Torsten Husén
(Chairman of IEA)

APPENDIX C

Statutes

INTERNATIONAL ASSOCIATION FOR THE EVALUATION OF EDUCATIONAL ACHIEVEMENT (IEA)

STATUTES

Name and aims

1. An international, non-profit-making, scientific association is constituted under the name of the "International Association for the Evaluation of Educational Achievement" (Hereafter known as IEA), the principal aims of which are:

   (a) to undertake educational research on an international scale;

   (b) to promote research aimed at examining educational problems common to many countries, in order to provide facts which can help in the ultimate improvement of educational systems;

   (c) to provide, within the framework of the Association, the means whereby research centers, which are members of the Association, can undertake cooperative projects.

Seat and legal status

2. (a) The seat of the Association is in Liege (Belgium). The Association is constituted in accordance with the Belgian law of the 25th October, 1919, regarding international non-profit-making scientific societies, modified by the law of the 6th December, 1954.

   (b) The co-ordination center of the Association is situated in Hamburg (Federal Republic of Germany). The co-ordinator is responsible for the day-to-day management of the Association.

Composition

3. Membership of IEA is restricted to institutions carrying out research in education (participating centers). Any such research center which has the status of a body corporate is eligible for membership, provided that it is sufficiently qualified and equipped with the resources required to undertake the research projects envisaged. Individuals working on the project may also be admitted as members. Membership is decided by the General Assembly (defined below). This decision requires a three quarters majority vote of the members of the General Assembly present.

4. Membership can be terminated:

   (a) by a three quarters majority vote of the members of the General Assembly present, where there are serious grounds;

   (b) by a vote of the same majority if a member ceases to satisfy the conditions of participation.

Those ceasing to be members forfeit all rights in the Association. They receive merely the amount of their paid-in capital and the fair market value of their contributions in kind, if any.

Organs and Administration

5. The organs of IEA are the General Assembly, the Standing Committee and the Bureau. The Standing Committee has its seat in Liege, Belgium.

6. The General Assembly, which comprises a representative from each participating center as well as individuals who may be admitted under paragraph 3, has complete authority. Each representative has one vote only. The General Assembly meets, in principal, once a year in September, and determines the general policy of the Association. It examines the reports and reviews the expenditure of the previous period and adopts the program and the budget for the following period. Each period lasts from 1st September to 31st August.

7. The General Assembly can establish such subsidiary organs as are necessary, either of a temporary or permanent nature. It determines, by a three quarters majority vote of all its members, the general composition, authority and term of office of these organs.

8. The General Assembly elects, by a three-quarters majority vote of its members present or represented, a Chairman.

9. The General Assembly elects, by a three-quarters majority vote of its members present or represented, a Standing Committee.

10. The Standing Committee comprises the Chairman and six other members, all elected by the General Assembly, which also has the power to dismiss them.

    The duration of the term of office of the Standing Committe members is normally three years. There should always be at least one Belgian member on the Standing Committee.

11. At the first election of the Standing Committee, one half of the members of that committee will be elected for a period of two years and the other half for a period of three years. Subsequent elections will be made for a period of three years.

    A member of the Standing Committee may be dismissed by the General Assembly, on the existence of sufficient grounds, by a three-quarters majority vote of its members.

12. The Standing Committee will be convened annually, or more frequently, at the discretion of the Chairman.

13. The functions of the Standing Committee are to make decisions, between meetings of the General Assembly, within the framework of the overall policy determined by the General Assembly; to appoint the Technical Director, the members of the sub-committees, staff and consultants, and to make decisions concerning expenditure within the limits of the budget determined by the General Assembly.

14. The Bureau comprises the Chairman and two members of the Standing Committee elected by that Committee.

15. The Bureau is responsible for the execution of decisions made by the General Assembly and the Standing Committee.

Authority

16. The General Assembly has the authority to make valid decisions when half of its members are present or represented. If the quorum is not reached, a second General Assembly, convened 12 weeks later, with the same agenda, may take decisions on the matters in hand with a three quarters majority of those present.

    Each member may register a proxy vote for another member, but only one other.

    All decisions made by the General Assembly, other than those stipulated in paragraphs 3, 4 and 25, require a simple majority vote of the members present.

17. The Standing Committee has the authority to make decisions when at least three of its members are present.

    All decisions made by the Standing Committee require a simple majority vote of the members present.

18. All documents involving the Association are, with the exception of special procurations, signed by the Chairman, who will not be required to justify his powers to any third party.

19. In all judicial actions undertaken by the Association, either as plaintiff or defendant, and in all prosecutions and lawsuits, the Association will be represented by the Chairman or one of the members of the Standing Committee appointed by the Chairman.

20. The resources of IEA comprise funds arising, for example, from gifts, grants, legacies, and profits from the sale of its publications. The General Assembly may also determine the amount of an annual contribution.

21. The IEA national costs will normally be secured by the respective research centers themselves.

22. IEA will finance its international costs by means of the resources listed in paragraph 20.

22a. In the event of profits being made by the Association, they may be used only for the aims stated in Article 1 of these Statutes. The members may take no share in the profits, and in their capacity as members may receive no gift, donation, grant or loan from the Association. The Association may not incur administration expenses foreign to the aims of the Association or pay disproportionate amounts (for example, salaries, allowances, honoraria, etc.) to any one person.

The auditing of accounts will be carried out each year by a competent international authority and the accounts are to be approved by the General Assembly.

Modifications and Dissolution

24. All proposals for modification of the statutes and for dissolution of the Association must come from one or more members of the General Assembly. All members of the General Assembly should be notified of such proposals (either by the Chairman, or in the absence of the Chairman by the other members of the Bureau) at least six months before the meeting of the General Assembly during which they are to be discussed. No decision on these matters can be taken without a three quarters majority vote of the entire General Assembly. If the quorum is not attained, a second General Assembly convened 12 weeks later with the same agenda, may take decisions on the matters in hand with a three quarters majority vote of the members present.

    In the case of dissolution, the General Assembly decides, by the same majority, on the steps to be taken for the liquidation of the Association and on the future of the assets remaining after the repayment of:

    (a) the paid-in capital, if any, and
    (b) the fair market value of the members' contributions in kind, if any.

    The remaining assets may only be used for aims in accordance with Article 1 of these Statutes.

    The decision of the Council regarding the use of the remaining assets requires the approval of the competent tax authority.

25. Any item which has not been provided for by the present statutes will be dealt with according to the provisions of Belgian law on the matter in question.

APPENDIX D

SCIENCE DOCUMENTS

APPENDIX D-1

FIRST GRID OF TOPIC AREAS AND OBJECTIVES IN SCIENCE

Subject Areas	Objectives															
	1		2		3		4		5		6		7		8	
	Obtaining Scientific Information		Interpreting Scientific Information		Theorisation Construction		Theorisation Utilisation		Comprehension		Application of Scientific Knowledge		Personal and Social Objectives		Philosophical Aspects	
	Present	Future	P	F	P	F	P	F	P	F	P	F	P	F	P	F
Earth Sciences																
1. Solar System																
2. Stellar Systems																
3. Meterology																
4. Earth's Crust																
5. Physical Geography and Geology																
6. Soil Sciences																
Biological Sciences																
7. Cell Structure and Function																
8. Homeostasis																
9. Cell Metabolism																
10. Cell Responses																
11. Concept of the Gene																
12. Diversity of Life																
13. Metabolism in Organisms																
14. Regulation in Organisms																

FIRST GRID OF TOPIC AREAS AND OBJECTIVES IN SCIENCE

Subject Areas	Objectives															
	1		2		3		4		5		6		7		8	
	Obtaining Scientific Information		Interpreting Scientific Information		Theorisation Construction		Theorisation Utilisation		Comprehension		Application of Scientific Knowledge		Personal and Social Objectives		Philosophical Aspects	
	Present	Future	P	F	P	F	P	F	P	F	P	F	P	F	P	F
15. Co-ordination and Behavior																
16. Reproduction and Development																
17. Human Biology																
18. Natural Environment																
19. Cycles in Nature																
20. Natural Groups and Their Segregation																
21. Population Genetics																
22. Evolution																
Physical Sciences: Chemistry																
23. Chemical Change																
24. Electro-chemistry																
25. Chemical Laws																
26. Chemical Processes																
27. Periodic System																
28. Energy Relationships in Chemical Systems																

FIRST GRID OF TOPIC AREAS AND OBJECTIVES IN SCIENCE

Subject Areas	Objectives															
	1		2		3		4		5		6		7		8	
	Obtaining Scientific Information		Interpreting Scientific Information		Theorisation Construction		Theorisation Utilisation		Comprehension		Application of Scientific Knowledge		Personal and Social Objectives		Philosophical Aspects	
	Present	Future	P	F	P	F	P	F	P	F	P	F	P	F	P	F
29. Rate of Reaction																
30. Raw Materials and Chemical Substances																
31. Chemical Structure																
32. Polymerisation and Polymers																
33. Chemistry of Life Processes																
34. Nuclear Chemistry																
Physical Sciences: Physics																
35. Measurement																
36. Time and Movement																
37. Forces																
38. Dynamics																
39. Energy and Machines																
40. Mechanics of Fluids																
41. Introductory Heat																
42. Change of State																

FIRST GRID OF TOPIC AREAS AND OBJECTIVES IN SCIENCE

Subject Areas	1 Obtaining Scientific Information		2 Interpreting Scientific Information		3 Theorisation Construction		4 Theorisation Utilisation		5 Comprehension		6 Application of Scientific Knowledge		7 Personal and Social Objectives		8 Philosophical Aspects	
	Present	Future	P	F	P	F	P	F	P	F	P	F	P	F	P	F
43. Kinetic Theory																
44. Light																
45. Vibrations																
46. Wave Phenomena																
47. Spectra																
48. Static Electricity																
49. Current Electricity																
50. Magnetism Electro-Magnetism and Alternating Currents																
51. Electronics																
52. Molecular and Atomic Physics																
53. Theoretical Physics																

APPENDIX D-2

TABLE 1

Expanded List of Teaching Objectives

Objective as it appears on the form	Explanatory expansion of the objective
1. Obtaining scientific information	1. The acquisition of scientific knowledge i) Observation of phenomena and its description in the pupils' own language. ii) Growth of vocabulary; scientific terminology as giving precision to scientific description. iii) Quantification: measurement as an extension of description: the selection and use of instruments; the appropriateness of measures: estimations and limits of accuracy. iv) Experimentation including the recognition of a problem, formulation of working hypotheses, selection of suitable methods and the devising of appropriate experimental procedures.
2. Interpreting scientific information	2. The organisation, interpretation and use of scientific knowledge i) Organisation: classification and processing of data and the recognition of relationships. ii) Presentation of data. iii) Generalisation, which includes extension of terminology in the form of definitions: functional relationships and their expression as scientific laws.

3. Theorisation - construction	3. The building up, formulation, use and refinement of theoretical models (mental pictures or concepts) i) The recognition by the pupil of the need for a theoretical model in order to relate different phenomena. ii) The recognition and appreciation of the relationships that are satisfied by the model. iii) The ability on the part of the pupil to make rational predictions following the use of the model.
4. Theorisation - utilisation	4. The use of theoretical considerations to direct observation and experiment. Evaluation and interpretation of experimental results at the theoretical level.
5. Comprehension	5. The ability to identify, comprehend, and recognise the significance of scientific knowledge in a new context whether it be obtained as the result of direct experience or derived from secondary sources.
6. Application of scientific knowledge	6. The application of scientific knowledge in a new and wider context and across subject boundaries.
7. Personal and social objectives	7. The development of personal skills, of hand and mind, and the encouragement of desirable attitudes and habits. Science in its historical perspectives. Technical and economic aspects of science: the relationships between scientific progress, technical achievement and economic development. Contemporary social aspects of science at the personal, communal, national and international levels, and especially the idea of human responsibility.
8. Philosophical aspects	8. The limitations of scientific explanation; Scientific methodology. Influence of the philosophy of science upon general philosophy.

List of Subject Areas. Table 2

Subject Area	Elementary 10+	Intermediate 14+	Secondary 16 and 19
Earth Science			
1. Solar System	Earth and moon in relation to the sun. Direction. Seasons.	Solar system. Explanation of apparent solar motions.	Planetary motion. Kepler's Laws. Newton's explanation.
2. Stellar Systems	The sky at night, constellations.	Stars and galaxies. Stellar distances and sizes.	Cosmology.
3. Meteorology	Weather	Weather maps and their interpretation.	Meteorology
4. Earth's Crust	Study of rocks and minerals	The Earth's crust as a source of materials and other resources.	-
5. Physical Geography and Geology	Weatherings. Fossils.	Earth forms, deposition, erosion. Fossils and fossilisation.	Geomorphology, stratigraphy, oceanography and palaeontology.
6. Social Sciences	Soil studies	Soils in relation to ecology and agriculture	Pedology
Biological Sciences			
Biology of the Cell 7. Cell Structure and Function	-	Organisms are made of cells. Cell structure and function.	Cell as unit of structure and function.
8. Homeostasis	-	Diffusion and Osmosis	Homeostasis-osmoregulation, permeability, membrane phenomena.

Subject Area	Elementary 10+	Intermediate 14+	Secondary 16 and 19
9. Cell Metabolism	-	Simple ideas of metabolism - Respiration, Photosynthesis.	Cell metabolism
10. Cell Responses	-	-	Regulation of cell response and cell behavior.
11. Concept of the Gene	-	The idea of an inheritable unit.	Gene and gene action.
Biology of the Organism			
12. Diversity of Life	The Variety of Life	Classification of plants and animals. Relationships between plants and animals.	Diversity of plant and animal form and its implications.
13. Metabolism in Organism	Ideas of breathing, digestion, etc.	Simple plant and animal physiology.	Metabolism in organisms and the structural adaptations involved.
14. Regulation in Organism	-	Regulation of temperature and water balance.	Homeostasis at the level of the multicellular organism.
15. Co-ordination and Behavior	Plant and animal reactions to external stimuli.	Plant and animal co-ordination and responses. Behavior.	Co-ordination. Regulation (nervous and hormonal). Behavior.
16. Reproduction and Development	Ideas of reproduction. Life histories.	Reproduction and Development. Metamorphoses.	Reproduction and Development.
17. Human Biology	Man as a living organism.	Man as a living organism.	Man in his physical and social environment.

Subject Area	Elementary 10+	Intermediate 14+	Secondary 16 and 19
Biology of Populations.			
18. Natural Environment	Plants and animals in natural environments.	Interrelationships between plants and animals and their environments.	Food cycles. Pyramid of numbers. Energy relationships in ecosystems.
19. Cycle in Nature	Simple food chains and food relationships.	Food chains: predators and scavengers.	Food cycles. Pyramid of numbers. Energy relationships in ecosystems.
20. Natural Groups and their Segregation	-	-	The concept of natural groups. Speciation. Modern taxonomy.
21. Population Genetics	-	-	Population genetics.
22. Evolution	-	Simple ideas of evolution.	Variation. Competition. Adaptation. Natural Selection.
Physical Sciences Chemistry Reactions			
23. Chemical Change	Bangs and smells	Chemical Change. Oxydation and reduction.	Chemical Change.
24. Electro-Chemistry	-	Electrolysis	Electro-chemistry. Redox reaction.
25. Chemical Laws	-	Conservation of mass.	Laws of chemical combination. Stoichiometry.

Subject Area	Elementary 10+	Intermediate 14+	Secondary 16 and 19
26. Chemical Processes	-	Laboratory and industrial processes.	Laboratory and industrial processes
27. Periodic System	-	Periodic table.	Periodic system.
Energetics			
28. Energy Relationships in Chemical Systems	-	Exothermal and endothermal reactions.	Energy relationships in chemical systems.
29. Rate of Reactions	-	-	Chemical kinetics.
Materials and Chemical Structure			
30. Raw Materials and Chemical Substances	Raw materials and their uses.	Raw materials. Purification and separation.	-
31. Chemical Structure	-	Elements and compounds ions, atoms, and molecules.	Chemical bonding and chemical structure. Modern atomic theories (models).
32. Polymerisation and Polymers	-	-	Polymerisation and polymers. Natural and synthetic Processes.
33. Chemistry of Life Process	-	Chemistry of respiration, nutrition and photosynthesis.	Bio-chemistry.
34. Nuclear Chemistry	-	-	Nuclear chemistry isotopes and radioactivity.

Subject Area	Elementary 10+	Intermediate 14+	Secondary 16 and 19
Physics Mechanics			
35. Measurement	Measuring and weighing.	Measurement. Physical units.	Physical measurement. Theoretical aspects. Errors and dimensional analysis.
36. Time and Movement	Timing	Motion. Velocity.	Kinematics. Vectors.
37. Forces	Pushing. Pulling. Twisting. Braking.	Force. Mass. Weight. Friction.	Statics. Equilibrium.
38. Dynamics	-	-	Inertia. Momentum. Gravitation.
39. Energy and Machines	Levers and pulleys	Levers, moments, Work. Energy.	Mechanical energy. Work. P.E. K.E. Machines.
40. Mechanics of Fluids	Floating and sinking.	Pressure. Upthrust in fluids. Flotation.	Hydrostatics. Hydrodynamics. Fluid flow.
Heat			
41. Introductory Heat	Expanding and Contracting. Thermometers.	Expansion. Thermometers. Transference of heat.	Quantitative heat; Specific heat. Expansion. Coefficients.
42. Change of State.	Melting and boiling.	Change of state.	Quantitative heat. Latent heats.
43. Kinetic Theory	-	Elementary kinetic theory.	Kinetic theory - gas laws. Thermodynamics: works and heat.
Light			
44. Light	Mirrors and lenses	Mirrors and lenses	Geometrical optics. Optical instruments. Photometry.

Subject Area	Elementary 10+	Intermediate 14+	Secondary 16 and 19
Waves and Vibrations			
45. Vibrations and Sound	Notes and noises	Sound, Instruments	Mechanical vibration. Longitudinal waves. Sound and accoustics.
46. Wave Phenomena	-	-	Reflection, Refraction. Interference. Diffraction. Polarisation.
47. Spectra	Colours	Colour. Spectra	Electromagnetic spectrum.
Electricity			
48. Static Electricity	Charges and sparks.	Static electricity.	Electrostatics.
49. Current Electricity	Batteries and bulbs.	Current electricity. Circuits. Units.	Direct current. Circuitry. Electrolysis.
50. Magnetism. Electro-magnetism and Alternating Currents	Magnets and Compasses.	Magnetism. Terrestrial magnetism. Electromagnetism.	Electromagnetism. Electromagnetic induction. A/C Transformers.
51. Electronics	-	Valves in circuits.	Electronics-thermionics, photoemission, semiconductors.
52. Molecular and Atomic Physics	Boundaries and bubbles.	Properties of matter.	Structure of molecular systems and properties of matter. Nuclear physics. Radioactivity.
53. Theoretical physics.	-	-	Relativity, wave mechanics.

APPENDIX D-3

MODIFICATION OF GRID OF TOPICS AND OBJECTIVES

Letter from International Committee to National Committees, May 1967

Completed grids have been received from ten National Centres together, in most cases, with comments of great interest and value.

Bulletin No. 2 Phase II has also been issued; it redefines the target populations (pages 2 and 3) and sets out the testing procedures (pages 4 and 5). These sections should be studied before any action is taken as a result of this report and in accordance with the timetable given in IEA/SCI/1, April-September, 1967.

Much of the success of the Project, so far as science is concerned, rests upon the production of suitable test items by all the contributing countries. This must be our main and urgent task during the next three months. In order to save time and effort, lists are attached of subject areas which are common to all, or nearly all, the participating countries for the groups who will be tested. These have been prepared after an analysis of the grids by the Chairman of the International Committee on Science and without prior consultation with other members of the Committee. This has been done in order to gain precious time and it is hoped that the concentration of effective effort that can result from this rather undemocratic action will justify it and offset any imperfections revealed in the lists.

It can be said that the grids have served their purpose in spite of their obvious weakness about which the International Committee were only too well aware. The purpose was to discover common subject areas on which to base the testing instruments, and to gain some impression of the objectives science teachers sought to achieve so that attainment could be assessed against these objectives.

It was apparent when the grids were analysed that the form in which they were given was much better suited to some countries than to others. Not unexpectedly the shortened list of teaching objectives came in for some criticism, as did the way in which the subject areas were expressed and arranged. Some contributors thought the list of objectives was too condensed so that discrimination was lacking, while others thought it was too long because teachers were not usually aware of the objectives they aimed to reach in their teaching.

The subject areas given in the attached lists should prove sufficient for the construction of the tests as there was a fair measure of agreement in the grids. The same cannot be said of the objectives, in which very little consistency could be detected even in apparently fairly obvious cases. It would seem wise therefore to adopt the following procedure in the construction of test items:

1. To devise test items on the subject areas contained in the attached lists and designed to test, as far as possible, the achievement of specified objectives.

2. To submit these test items to groups of experienced science teachers, asking them to specify what objectives in their opinion the items assess.

3. To send the test items to IEA Hamburg, indicating the target population for which they are intended and the objectives they are thought to assess. For the time being, it might be wise to use the objectives as amplified in Table 1 IEA/SCI/3 as a basis, but it is to be hoped that National Committees for Science will feel free to enlarge and modify that list, which will be revised later in the light of the suggestions and comments which have been, and will be, received.

4. In general, objective type tests which can be marked automatically are looked for and the International Committee will be particularly pleased to receive test items of a new kind designed to assess specific scientific abilities, but this should not inhibit the production of items of a subjective character so long as these are felt to be indispensable to the proper assessment of a given ability.

5. The examination of the pre-tests by the National Committees during the period October/November 1967 should ensure that the testing instruments cover an adequate sample of the subject areas, measured against the teaching objectives, for satisfactory results to be obtained. But, as was stated earlier, the success of the whole operation depends to a very high degree upon the skill, energy and ingenuity displayed by the National Committees working in conjunction with a wide circle of practising science teachers. The active co-operation of National Committees is therefore earnestly sought.

IEA

SCIENCE

List of Subject Areas Suggested for the Production of Test Items

Target Population (1) 10+ (see Bulletin 2 pages 2 and 3)

Topic	Ref. in Grid	Comment
Earth Science		The general impression gained from the grids and accompanying comments is that at this stage earth science is taught as part of the general exploration of the natural environment, often as geography. Test items should be constructed with this informal approach in mind. The growth of a more precise vocabulary to describe familiar phenomena would be a legitimate objective to test. Most countries will not have introduced experimental science in anything but a very restricted sense.
The Earth in the Solar System	1	
Weather Studies	3	
Physical Geography	5	
Biological Sciences		Although human biology as such is not often taught, the growth of biological ideas is generally seen in relation to human life and man's activities.
Diversity of Life	12	
Breathing, feeding and the elimination of waste	13	
Reproduction	16	
Natural Environment	18	

Topic	Ref. in Grid	Comment
Chemistry		Few countries give any substantial attention to chemistry at this stage and yet important chemical ideas, e.g., of chemical substance, extraction, separation, purification, and chemical properties must be emerging, so that items might be set under these two topic headings.
Chemical Change	23	
Raw Materials	30	
Physics		Measurement, although perhaps a more obvious part in physics, is misplaced here and should be viewed in a more general way as applying to all the sciences.
Measurement	35 (see comment)	
Motion	36	
Forces and their applications	37/38	
Floating and Sinking	40	
Expansion and Contraction	41/42	
Melting and Boiling		
Mirrors and Lenses	44	
Sound especially Musical Instruments	45	
Elementary Electricity	48/49/50	

Target Populations (2 & 3) (see Bulletin 2 pages 2 and 3)

Topic	Ref. in Grid	Comment
		It appears that there is a fair amount of common ground at these levels but there are indications that the depth of treatment and the extent of the practical work vary considerably. It is suggested that test items should be based on a quantitative treatment using simple numerical examples when they serve to illustrate fundamental principles and to establish relationships more clearly. Qualitative relationships and test items are not, however, to be omitted entirely.
Biological Sciences		
The cell as a unit of structure and function (simple outline only)	7/9/10	
Diversity of life and inter-relationships	12	
Biology of selected organisms to illustrate characteristics of life	13/14/15/16	With some reference to human beings. Relation of structure to function. Adaptations to mode of life. Elementary anatomy and physiology of representative plants and animals.
National environment and seasonal rhythms	18	
Chemistry		Elementary chemistry can be tested at this stage, e.g., notions of classification elements, compounds, mixtures; simple reactions leading to activity series, acids, bases and salts, laboratory preparations of common substances, etc.
Chemical substances and processes	23/26/(28/29) 30/31	very simply

Topic	Ref. in Grid	Comment
Physics		Signs of change from traditional subject matter to more modern material are more obvious here than in biology and chemistry. To meet this situation perhaps we ought to collect items reflecting traditional treatments and modern ones as well. For this reason subject areas not entirely common to all countries have been included exceptionally.
Measurement	35	Especially elementary ideas of accuracy, approximation; fundamental and derived units.
Motion (Kinematics)	36	
Forces (equilibrium, causing motion)	37	
Energy (changes and conservation)		
Machines (mechanical advantage, efficiency)	39	
Mechanics of fluids (Archimedes Principle Buoyancy)	40	
Elementary heat (expansion, change of state)	41/42	
Light (elementary optics)	44	
Sound (vibrations and waves simply)	45	
Elementary Electricity	48/49/50	Simple circuits. Effects of electric currents (magnetic, heating(chemical), that is qualitative treatment only, not Ohm's law.

Target Population (4) (see Bulletin 2 pages 2 and 3)

Topic	Ref. in Grid	Comment
		Practice with regard to the earth science is so variable that as with Populations 2 and 3 there is not sufficient common ground to justify testing.
Biological Sciences		
Cells as units of structure and function (in more detail)	7/8/ 7/8/ 9/10	
Genetics, practical and theoretical aspects	11	
Diversity of life, incl. classification	12	
Biology of organisms (in more detail)	13/14/ 15/16	
Natural groups, speciation and classification	20	
Evolution	22	
Chemistry		
Basic facts, methods, principles and theory of chemical sciences	23/24/ 25/26/ 27/28/ 29/30/ 31	
Chemistry of carbon compounds and of the life processes	33	Practice varies considerably from country to country but it appears desirable at least at this preliminary stage of the project to produce some questions on organic chemistry.

Topic	Ref. in Grid	Comment
Physics		See general comment on physics for Populations 2 and 3.
Physical measurements, constants, methods of measuring, theoretical aspects	35	
Statics and Dynamics	36/37/38/39/40	More emphasis should be placed upon the experimental and applied aspects than upon the mathematical ones.
Heat	41/42	
Kinetic theory	43	
Light	44	
Sound and wave phenomena in general	45/46	
Spectra	47	
Electricity and Magnetism	48/49/50	Practical and theoretical aspects including quantitative work
Electronics	51	
Atomic physics	52	

APPENDIX E

READING COMPREHENSION RATIONALE

APPENDIX E

READING COMPREHENSION RATIONALE

Since reading comprehension is an important tool for the understanding of most other subjects, the IEA Council has decided to include it in Phase II of IEA.

The Reading Comprehension Committee considered a number of views of reading comprehension and concluded that reading comprehension regarded as instrumental to learning is common across countries. Because of the difficulty of handling aesthetic and affective characteristics of written material in translation, and because there is a literature committee which will include these in their study, the Committee thought it best to limit reading comprehension to cognitive aspects.

It seemed advisable to the Committee to try to reproduce in the tests the commonest situations in which reading is used in schools. The three main situations appear to be:

a. Assignment type reading where questions are answered with the passage available for reference (III-1-a).

b. Reading to remember and recall (III-1-b).

c. Directed type reading, in which particular lines of inquiry are prescribed, and questions are answered without the passage available (III-1-c).

For the same reason, the Committee thought it best to use paragraphs rather than isolated sentences to test reading comprehension. Oral reading was ruled out because it would necessitate individual testing.

A number of related measures could have been used but those chosen were thought most likely to cover the dimensions of variation which contribute to reading comprehension. In this respect, vocabulary seemed an obvious choice, but the problem of finding words of comparable levels of difficulty from country to country argued in favor of a measure not so dependent on word familiarity. A test of word meaning was therefore adopted which depends on judgment between pairs of words where the difficulty lies in their relationship rather than in their rarity. (III-2-a)

The intimate relationship between rate and comprehension persuaded the committee that a separate measure of speed of reading was required. A device is to be used in which rate will be measured in relation to a constant level of difficulty. (III-2-b)

Individual differences in adjustment of rate to the level of difficulty suggested a further test in which the degree of adaptation of speed of reading to level of reading difficulty could be assessed. (III-2-c)

A study of reading comprehension aimed at selected school populations (II-Target Populations) should enable us to test a number of hypotheses. The Committee has drawn up a list of suggested hypotheses which is not meant to be exhaustive. Comments, and further suggestions are asked for: when making any comments or suggestions, however, it should be borne in mind that comparisons across countries are of primary interest in such a survey.

I. PROPOSED HYPOTHESES

1. Internal structure of the test battery

a. What are the relationships between the different tests of reading comprehension in the different countries and how do these relationships change according to the level of education of the target populations?

b. What are the relationships between reading comprehension, rate of reading, vocabulary, adjustment of rate of reading, in the different countries; how do they change according to the level of education of the target populations?

2. Educational practices and background

The level of reading comprehension may be influenced by:

a. methods of instruction at the beginning of reading (i.e., analytic, synthetic, mixed).

b. pre-school instruction in reading.

c. importance of reading in the curriculum (definite reading lessons).

d. formal training in grammar, in vocabulary.

e. kinds of reading assignments if any (literary appreciation, extensive reading, memorizing poems or texts, etc.).

f. number of teachers the child has each week.

g. type and size of school.

h. type of studies (i.e., general, technical . . .)

i. specific subjects studied.

j. grade.

k. advance or retardation in studies.

l. the grade at which a second or foreign language is introduced.

3. It may also be influenced by the availability of reading material.

a. in the school (does the child have textbooks for different subjects, is there a school library?)

b. in the home (books, newspapers, magazines)

c. in the community (can books, newspapers, magazines be bought or borrowed?)

4. The environment outside the school may have an influence on reading comprehension.

a. Occupation and social standing of the parent(s)

b. Level of instruction of the parent(s)

c. Amount of exposure to mass media (radio, TV, films)

d. Is the language of instruction the child's first language?

e. Is the child monolingual?

5. The interests of the child and his attitudes may have an influence on reading comprehension.

a. Preference for some of the subject-matters (by direct ranking, by ranking of the different texts of the test, etc.)

b. Interests in leisure activities

c. Attitudes toward reading (child's idea of the aims of reading, why is it important to be able to read, etc.)

d. Ambitions and aspirations of the child

e. Attitudes of the child toward school (does he regard it as a compulsion)

6. Study and reading habits of the child may influence his reading comprehension.

(e.g., finger tracing, vestigial oral reading, etc.)

7. There probably is a relationship between the child's achievement in different subjects and/or overall achievement and his level of reading comprehension.

8. Physical defects might impair the child's reading.
(e.g., eye sight, hearing, speech defects)

II. TARGET POPULATIONS

1. The most important populations seem to be at:

a. Towards the end of primary education (around 10 or 11): This stage was considered to be important in the testing of reading comprehension for several reasons. One is that in many countries almost all the children are still in school and have followed the same course of instruction. Another reason is that it is the stage at which, in most countries, children have finished formal training in the mechanics of reading. It is also the stage when children, entering the secondary school, are expected to be able to read more independently.

b. The pre-university level: This is the best stage at which to attempt to assess the overall effect of primary and secondary education. At this stage most pupils probably have completed any formal instruction or training they will get in reading. In the university, they are expected to use extensive reading as one of the main tools of study.

2. Populations also considered of interest in such a survey at a third priority level but which would be most difficult to identify clearly and consistently across countries are those at:

a. the level of transition from general to specialized education (about 14-15) and/or

b. the terminal point of education for many pupils which usually coincides with the end of the compulsory education period.

III. KINDS OF MEASUREMENT ENVISAGED

As it has already been stated, it is intended that the reading comprehension tested will be limited to the communication of ideas and will exclude perception of feeling, appreciation of aesthetic quality or purely literary devices. Other aspects of comprehension which are excluded are:

a. the comprehension of other symbolic media, e.g., plans, diagrams, charts, graphs, maps, etc.

b. listening or aural comprehension - one exception might be a short test to be sure that pupils especially at lower levels have understood what they are to do (i.e., what the task is, how they are to mark the answer sheet, etc.). Any testing of purely oral reading is also excluded.

1. In view of the above considerations, the following types of reading comprehension tests are envisaged:

a. Silent reading of passages continuously available to the pupils followed by written questions.

b. Silent study of a passage which is then removed before the questions are asked.

c. Silent study of a passage after the pupil has been given some direction as to the type of information he is expected to get from the passage. The passage is to be removed before the questions are asked.

N.B. The questions on the passage will be multiple choice type and/or completion type with no choices given (the questions being so designed that there is only one right answer).

2. Supplementary tests

The aim of these is to test diagnostically some of component skills of reading comprehension.

a. A word meaning test

b. A separate measure of rate of reading is envisaged. It will be carried out at a simple comprehension level.

c. A test of adjustment of the speed of reading to the difficulty of the material.

APPENDIX F

LITERATURE DOCUMENTS

APPENDIX F-1

THE ELEMENTS OF CRITICISM AND INTERPRETATION[1]

Alan C. Purves

For a study of the ways in which students, teachers, and critics approach a literary work, we have found it necessary to devise some means of describing the critical process.[2] The usual classifications - formalist, Aristotelian, Marxist, Freudian - are not adequate because such terms do not really show what happens in a piece of writing, and because few essays fall neatly into any one niche (and they should not, for were they to fall so, the end of criticism as a human encounter with literature would be at hand). One can more clearly perform the task by identifying the intellectual operations that constitute a piece of critical and interpretive writing.

These operations would be discrete, elements, as it were, of criticism and interpretation, which may be combined in many ways to form a critical method. In fact, the experienced reader might employ nearly all of the lements - at least subconsciously - every time he confronts a new work. Some he ignores, some he emphasizes, some he subordinates to others, as he goes about the act of reading and writing about his reading.

As one looks at the list below, one must realize that within the four categories - or between them, for that matter - there is no rigid hierarchy. Certain sub-groupings represent the idiosyncracy of the writer, yet the terms themselves, as well as their placement within categories, represent the consensus of our advisers, and, I hope, are neutral. What value one is to assign to the use of any element or combination of elements is beyond the purview of this study.

---

1. A brief version of this paper appeared as an interim report of The Horace Mann-Lincoln Institute of School Experimentation, New York 1965.

2. This study is a project of The Horace Mann-Lincoln Institute of Teachers College, directed by A. W. Foshay. We are grateful for the interest, assistance, and criticism of Father Walter Ong, S.J., Professors Josephine Miles, Marshall McLuhan, Wilbur Scott, Lewis Leary, Robert Gorham Davis, Barry Ulanov, Stanley Edgar Hyman, Richard P. Adams, Arlin Turner, Albert Hofstadter among many others. We are also grateful to the countless teachers and students whose work provided a constant check on ours.

Copyright, ©, 1965
Teachers College, Columbia University

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY A. W. Foshay, Prof. of Educ. Teach. Coll., Col. Univ.
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

iii Summary of Critical and Interpretive Modes

- I. Engagement - Involvement
  - A. Reaction to the content of the work
    - 1. Moral reaction
    - 2. Word associations
  - B. Identification of reader with the work
    - 1. Relation of incidents to those in the reader's life
  - C. Redaction of work from a point of view different from that stated
  - D. The impressionistic monologue which speaks more of the reader's sensibility than of the work
  - E. The divergent response

- II. Perception of the Work
  - A. Analysis of the pattern of order of the word
    - 1. Language
      - a. Grammar and typography
      - b. Syntactic and rhythmic patterns
      - c. Diction
    - 2. Literary devices
      - a. Rhetorical devices
      - b. Imagery or image patterns
      - c. Larger literary devices
    - 3. Content and structure
      - a. Narrative time or action
      - b. Plot structural analysis
      - c. Character analysis
      - d. Character relationships
      - e. Setting
    - 4. Tone or mood
      - a. Establishment of the writer's or speaker's tone or point of view on the mood of the work
      - b. Pace of the work
      - c. Mask
    - 5. Total structure or Gestalt of the work
    - 6. Rhetoric of the work
      - a. Relation of technique to meaning or effect
      - b. Presentational elements (drama)
    - 7. Typological figures
      - a. Common symbols
      - b. Allegorical relationships
  - B. Classification of the work
    - 1. Generic classification of any part
    - 2. Traditional
    - 3. Contextual
      - a. With the author's other works
      - b. Biographical
      - c. Intentional
      - d. Historical
      - e. Intellectual

- III. Interpretation of the Work
  - A. Interpretation of parts (may follow any mode in B below)

    1. Interpretation of a passage of the work as a key to the whole work
    2. Derivation of specific symbols
    3. Conjecture about the past and present in the work
  B. Interpretation of the whole
    1. Mimetic
      a. Social
      b. Political
      c. Historical
      d. Moral
      e. Psychological
    2. Typological
      a. Psychological
      b. Theological
      c. Political
      d. Philosophical
      e. Archtypal
      f. Aesthetic
    3. Moral
      a. Ethical
      b. Poltical or governmental
      c. Social or sociological
      d. Psychological

IV. Evaluation of the Work or any Part of the Work

  A. Affective
  B. Formal
  C. Coherence
  D. Generic
  E. Traditional
  F. Intentional
  G. Mimetic plausibility
  H. Thematic Plausibility
  I. Symbolic Plausibility
  J. Moral acceptability
  K. Multifariousness
  L. Rhetorical

iii The Categories (Engagement - Involvement, Perception, Interpretation, Evaluation)

The establishment of these categories was the major work of our advisers, and represents a general picture of the intellectual operations one performs with literature. Engagement-Involvement is the best term we could find to define the combined acts of entering into a literary work and experiencing it or surrendering to it. Alternative terms like response or assimilation seem over-restrictive. To some extent, engagement-involvement is of a different order from the other three categories in that it encompasses a process and in that it is often not recognized as a part of the critical tool-kit. Yet some excellent criticism has arisen from engagement-involvement, and although it has fallen from favor among many critics it appears a necessary adjustment to the reading of literature and is certainly not to be discussed as irrelevant. At times, in fact, one is hard put to distinguish a statement that relates a work to the reader's life from an interpretation of the work (the difference may simply be one of generalization).

The second category, perception, is almost self-explanatory: it encompasses both those parts of a work that a person looks at and the lens through which he looks at them. The first we might call an inductive mode of vision, the second deductive; the first arises from a point-of-view that comprehension comes from a regard to detail, the second from a regard to larger structure, a pattern of literature rather than of the literary work.

Interpretation, not unlike engagement-involvement, defines that area in which the work and the readers interact. It is the act by which the reader, who has established the "otherness" of the work, seeks to connect it with himself and his world. He says what the work means, which is itself a generalization, and then proceeds to put it in an even more general scheme. The process is one of abstraction or analogy and arises from a perception of the work as a whole that exists in a larger, non-literary universe. At a simpler level, it is also the drawing of inferences from the work, or from a part of the work. Interpretation ranges from the simplest conjecture about the future beyond the work through an attempt to see the work in relation to the known world to an attempt to assimilate the work into a larger moral order. Each mode of interpretation then arises from a particular view of the function of literature in society: as an imitation of life, as a distillation of reality, or as a medium of judgment about life.

Evaluation, the last category, encompasses the process of judging the work by some criterion, personal or objective. In this category, the order of elements roughly parallels the order of the elements in the other categories. One's judgment of a work follows from one's engagement-involvement, perception, or interpretation of the work; there would seem to be no other premise of taste.

These four categories exist as the general framework of the elements, but they exist in no particular order or procedure. Any one way precedes any other in time or in logic: one's reaction may follow from or produce one's interpretation or perception; an evaluation may be instantaneous or the result of a rigid examination of the text. In fact, an attempt to impose any rigid structure on these interwoven mental acts would be useless, because we do not know and can probably never know just how the mind of a reader works.

Before proceeding to define each of the elements, I should like to make several points about them. First, some of the words or phrases listed are not in themselves elements, but simply terms used to group elements: i.e. II A (Analysis of the pattern or order of the work);

II A 1 (Language); II A 2 (Literary devices); II A 3 (Content and Structure); II A 4 (Tone or mood); II A 6 (Rhetoric of the work); II A 7 (Typological figures); III A (Interpretation of parts): and III B (Interpretation of the Whole). Second, if these elements are used as a basis for analyzing a piece of writing, some sentences would not be included - sentences which are included for rhetorical effect or sentences which cite criteria for literature but do not actually judge the work. Third, for convenience the elements of II A (Analysis) are listed by the thing perceived or analyzed, and those of valuation by the criteria. The last, and most important, point is that although many of the elements could be subdivided, I think such a task would be fruitless and would render such a classification fruitless. There could be subdivisions based on genre or on schools of interpretation, but they would tell us little more about the cast of the writer's mind. It is less important, I think, to know whether a critic's political interpretation is rightist or leftist than to know that the critic sees the work as the reflection of a political idea. If anyone wishes to atomize the elements, he is free to do so.

iv The Elements of Engagement-Involvement

The expression of each of these elements seems to reflect more of the reader than of the work. That such reflection is not unimportant to the literary experience, few would deny, nor would they that it is related by some alchemy to the other parts of the critical and interpretive act.

I A Reaction to the content of the work incorporates diversity: interest, boredom, horror, amusement.

I A 1 On another level, reaction may take the focus of such a moral judgment or conjecture about a character or event as seems to arise from a failure to distinguish between art and life.

I A 2 Word associations are a minor form of this reaction, whereby a word in a work provokes such a reaction from the reader that his attitude toward the text is admittedly prejudiced.

I B Identification of the reader with the work is close to what has been called the vicarious experience. A reader might say, "I felt as if I were there," or otherwise indicate his submission.

I B 1 Relation of incidents to those in the reader's life is a form of identification, which, in an essay, often leads to a discussion of the reader's life, not the work.

I C Redaction refers to the process whereby a reader second-guesses the author either by changing the point-of-view or rearranging some aspects of the work. Although the element might be associated with evaluation, a dissatisfaction with the work, it seems more clearly an off-shoot of a person's experience of the work and dissatisfaction with that experience.

I D The impressionistic monologue, not unlike I B, is the expression of the reader's personality more than of his view of the work. At times it stems from an attempt to transmute the work into a second work of art - the first seen through the lens of a heightened sensibility --; more often it embodies the reader's fixing upon an idea, event, or mood, divorcing it from the work, and expanding upon it.

I E The divergent response is an extreme form of I D. If someone were asked to discuss a poem and instead wrote another poem, or an essay on his personal problems, such a response would be divergent more than impressionistic.

v. The Elements of Perception

II A 1 a Grammar and typography include such phenomena as odd devices of punctuation, grammatical errors, or visual effects (the odd spacing or indentation of poetic lines).

II A 1 b Syntactic and rhythmic patterns include meter, cadence in prose, phonetic patterns, and other stylistic devices involving the ordering of language.

II A 1 c Diction encompasses the author's choice of words, whether level of usage, lexical oddity, or even semantic ambiguity.

II A 2 Although some might consider these elements just of language, many would separate them because they connect language and content and are not simply linguistic considerations. These elements are to be differentiated from II A 6, because the former remain at the level of a catalog.

II A 2 a Rhetorical devices include such traditional figures as metaphor, metonymy, and zeugma.

II A 2 b Imagery refers to those striking, repeated, or characteristic references to a particular milieu that an author may make.

II A 2 c Larger literary devices include diologue, description, narration, and others which are not generic but which constitute parts of a work.

II A 3 a Narrative line or action refers to the order of events as the author presents it. In an essay, this element is signalled by a repetition or paraphrase of the text.

II A 3 b Plot or structural analysis differs from 3a in that here the reader is dealing with the author's ordering of events rather than the events themselves; he perceives a pattern in the text.

II A 3 c Character analysis includes the attempt to explain motivations or otherwise describe a character. It remains at the level of perception as long as the reader does not conjecture about the character's motivation or past where these are not given.

II A 3 d Character relationships involves the discussion of static relationships - the positions that a character might have vis-a-vis another. As soon as the relationship becomes dynamic, the perception would be of plot.

II A 3 e Setting refers to the perception of the locale, either in itself or in relation to narrative or character.

II A 4 a Establishment of the author's or speaker's tone or the mood of the work may seem close to reaction, yet many critics would agree that despite the difficulties of describing tone and mood objectively, the attempt is to be objective. This element will often be supported by elements II A 1 or II A 2 and their appearance might well serve as evidence of objective intent.

II A 4 b Pace is even less easy to establish than tone; "the work moves quickly" is close to impressionism, yet the same criterion as for II 3 A a applies.

II A 4 c Mask involves II A 4 a, but assumes a duality in the work: that the narrator and the author are not the same - or that they are.

II A 5 Total structure or Gestalt of the work refers to an attempt to describe the whole work, often in terms of a spacial metaphor ("it is circular"; "it turns on itself") or even in terms of a musical metaphor ("a rondo form").

II A 6 a Rhetoric of the work: the relation of technique to meaning or effect is perhaps a step beyond simple perception; it is the attempt to relate perception to the meaning, to explain parts of the work in terms of the whole, or to relate minor with major aspects of the work. This activity is based on the reader's assumption that the work is organic, that everything in the work is designed to aid in creating the work's meaning and effect.

II A 6 b Presentational elements refers to the aspect of the rhetoric peculiar to drama, the relation of the visual to the aural. It follows from the reader's attempt to visualize the play.

II A 7 a Typological figures: common symbols identifies the perception of those common symbolic referents in literature that cannot be taken as anything but symbolic. The perception of these seems close to typological interpretation, but one can distinguish the two as follows: Blake's Lamb in "Little Lamb, who made thee," cannot but be identified with Christ and the Christian. His Tyger, however, is a tiger and although it can be interpreted as power, or Christ, or truth, it remains a tiger.

II A 7 a Allegorical relationships, like common symbols, refers to the reading of those words, like the beast fables or Pilgrim's Progress in which the allegorical level is readily available to any reader. Some works of Kafka, too, demand this sort of perception, even though the specific allegorical reference is debatable.

II B Classification of the work as a general category refers to those perceptual acts which see one work as a part of literary history, and as a specific category to those acts which see the work as related to a literary prototype.

II B 1 Generic classification of any part defines the acts of typing a character (the servant of Roman Comedy), of making analogies between aspects of works, or by typifying an aspect of the work.

II B 2 A traditional is like a genuine classification, differing only in that the work is seen as a point on a continuum within the genre not merely as an object to be put in a drawer. The reader may go on to show the work's deviation from the tradition.

II B 3 This group of elements involves the act of seeing the work as a product of its author and of a point in time. The act is in its simple form classificatory, but the reader may well go beyond that act to a discussion of the relation of the work to its matrix. The sub-elements list the predominant focus of contextual classification: with the rest of the author's work (a), with the author's life (b), with the author's expressed or inferred purpose (c), with the events at the time of the work's creation (d), and with the history of ideas or a philosophic or religious point-of-view associated with the author or his times (e).

vi The Elements of Interpretation

III A Interpretation of parts is the element of interpretation in which the reader singles out a particular segment of a work and proceeds to generalize about it, although he may make no generalization about the whole work.

III A 1 Interpretation of a passage as a key resembles III A, but the reader here asserts the passage to be a reduction of the whole work.

III A 2 Derivation of specific symbols includes that process by which a reader invests an object with typological significance (see the explanation of II A 7).

III A 3 Conjecture about the past and present of a work is the form of interpretation closest to I A (Reaction), but can be differentiated thus. Interpretative conjecture seeks to ascertain motive or thought within the time limit of the work, and from data given in the work. The conjecture of reaction seeks to continue the work beyond the limits given by the author - to wonder, for instance, whether Hamlet will go to heaven or whether a couple really will "live happily ever after."

III B 1 Mimetic interpretation, like the rest of the elements in III, refers to a mode of interpretation. This mode imputes a reader's seeing the work as a mirror of the world, either generally or in one of the five specific ways. The reader says, "This is the way the world is." If the work is an heterocosm, it is one which is closely connected to the world around the reader.

III B 1 a Social mimetic interpretation refers to a world seen as the interaction of types of people, classes, or societies.

III B 1 b Political mimetic interpretation refers to a world seen as a power structure made up of the governors and the governed.

III B 1 c Historical mimetic interpretation refers to a world seen at a specific time in the past or at recurring similar times (the fall of empires).

III B 1 d Moral mimetic interpretation refers to a world seen as made up of ethical forces.

III B 1 e Psychological mimetic interpretation refers to a world made up of conflicts within the mind of individual man.

III C 2 Typological interpretation sees the work not as a mirror, but as an abstract or a highly generalized pattern of the world; the work becomes at least typical, at most allegorical or a metaphor of some pattern of existence. The patterns may be psychological (a), theological (b), political (c), philosophical - a systematic philosophy - (d), archtypal - a pattern of repeated figures or movements like fertility, harvest, sacrifice - (e), or aesthetic - a pattern in which all art is about the artist.

III C 3 Moral interpretation sees the work as a proclamation about how the world should be; it sees the author as hortatory - either overtly or covertly so. The author is seen as taking a stand, either ethical (a), political (b), social (c), or psychological (d).

vii The Elements of Evaluation

These elements are cast in the form of criteria, either for a subjective or objective appraisal of the work.

IV A Affective evaluation uses the criterion of emotional appeal. Either the work succeeds or fails in moving the reader, or in presenting its dominant emotion with sufficient intensity.

IV B Formal evaluation uses the criterion of aesthetic evaluation. The work may or may not fulfill its function, succeed or not to use all of its parts coherently like a good ballet. Often the untrained reader will say that he does not "like the looks" of the work; this, too, expresses a formal criterion, albeit weakly.

IV C Rhetorical evaluation uses the criterion of effective use of form to present and enhance content. It judges work by adequacy of
5 parts to the whole.

IV D Generic evaluation uses the criterion in which the reader uses the abstract notion of genre to judge a work. The best example is the critic who says, "It's a bad form; it does not rhyme."

IV E Traditional evaluation uses a criterion akin to that of genre, but less rigid; it judges the work according to the history of its type in form or content. Its criterion, then, is flexible, since tradition is, within limits, continually modified.

IV F Intentional evaluation uses the criterion of the author's expressed or inferred intention.

IV G Mimetic plausibility is the criterion of surface veracity, or, on a more rarified level, of the adequacy of the author's vision of life when compared with the reader's vision.

IV H Thematic plausibility is the criterion of sufficiency of meaning. It asks of the work that it have an import equal to the reader's set of values. It differs from IV G in that it asks, "Does the author think of life as I do (or as most do)," and IV G asks, "Does the author represent a world in which I can believe?'

IV I Symbolic plausibility is the criterion of congruence of patterns. It asks of the work that its abstraction of the world accord with the reader's abstraction. This criterion might also be one of the rhetoric of symbols: the reader might see the work as reaching for a symbolic structure and succeeding or failing in achieving it.

IV J Moral acceptability is the criterion of lessons. "Does the work teach what I consider morally correct?" asks the reader.

IV K Multifariousness is the criterion of levels. The reader asks of the work that it be interpreted in many ways and judges accordingly.

viii Conclusion

This schema, I hope, will be of use to critics and teachers. By separating the elements of criticism and interpretation, one can possibly regroup them; one can observe what it is he or his students do and seek to rectify omissions; one can even see the relationships which exist between certain works or genres and certain elements. Finally, the literary historian can use the schema to examine the critical temper of a past age and, perhaps, can draw inferences about the characteristic ways of approaching reality that are indicated by the characteristic ways of approaching the art work. It is this purpose more than any other that provoked the schema; yet a tool of intellectual history may become a tool of education.

Appendix B KEY FOR SCORING

I. ENGAGEMENT

01 Rhetorical filler
02 Reaction to content
03 Word associations
04 Relation of incidents
05 Identification with work
06 Redaction of work
07 Impressionistic monologue
08 Divergent response

II. PERCEPTION

09 Narrative line
10 Plot analysis
11 Character analysis
12 Character relationships
13 Setting
14 Language
15 Diction
16 Literary or rhetorical devices
17 Imagery
18 Tone
19 Mask, point of view
20 Total structure (Gestalt)
21 Rhetoric (Form & content)
22 Presentational (Drama)
23 Classification of work (as whole)
34 Generic (part of work)
25 Traditional
26 with author's work
27 Biographical
28 Intentional
29 Historical
30 Intellectual
31 Common symbols
32 Symbolic relationships
65 Typographical-grammatical
66 Reading comprehension
67 Perceptual general

III. INTERPRETATION

33 Passage as key
34 Derivation of symbols
35 Mimetic
36 Mimetic-social
37 -Political
38 -Historical
39 -Moral
40 -Psychological
41 Typological
42 Typological-Psychological
43 -Theological
44 -Political
45 -Philosophical
46 -Archetypal
47 -Aesthetic
48 Moral-Ethical
49 -Political
50 -Social
51 -Psychological
68 Interpretive General
69 Interpretation of parts
70 Conjecture

IV. EVALUATION

52 Affective
53 Formal
54 Generic
55 Traditional
56 Intentional
57 Mimetic plausibility
58 Thematic plausibility
59 Symbolic plausibility
60 Moral acceptability
61 Multifariousness
62 Rhetorical
63 Unspecified
71 Citation of Criteria (Art)

V. Miscellaneous

64 Unclassifiable general

Numbers 65, 69, and 70 are sub-categories of the elements that appear frequently enough to become categories in their own right. Numbers 01, 63, 64, 66, 67, 68, and 71 were added to classify sentences that appear in many essays, but that are not elements of criticism or interpretation. It is expected that as the study progresses certain elements will receive greater or less emphasis and that certain elements will be collapsed.

Appendix C. STATUS OF PROJECT

As of the summer of 1965, an outline of the elements of criticism and interpretation has been developed, together with a scoring sheet for student essays.

Sample essays by approximately 100 American, 50 Belgian, and 50 German students, age 13 and 17, all writing on a short story have been gathered. The essays will be scored and the data analyzed by the end of the summer. This work should test inter-reader reliability and should give preliminary data to give tentative support to Hypothesis II and III a (the major hypothesis of the project).

In addition, preliminary work on the analysis and comparison of the oral and written responses by a small group of students to a variety of literary works will have been completed.

An anlytic scheme has been developed in cooperation with Teachers College Computer Center. Results of a first analysis of the data obtained from the papers above will be presented to the participants in the IEA meeting at Hamburg, August 30, 1965.

APPENDIX F-2

LITERATURE RATIONALE

The general purpose of this study is to shed light on the relation between the humanistic aims of literary education, upon which we believe there would be broad agreement, and both school policies and practices and the place of literature in the cultures of different nations.

This we propose to do by examining the performance in, and attitudes towards, literature of students at the end of compulsory education and at the end of pre-university training.

The vital educational "unit" in literature teaching might be said to be the individual student's reaction to an individual work of literature, since, however, this is an imponderable, this study will focus upon his expressed, or formulated, responses. We expect that there will be characteristic national and cross-national patterns in such responses and that these will be related to aspects of the student's literary education.

HYPOTHESIS I

Characteristic patterns of expressed response will be closely related to knowledge about and attitudes towards literature.

There are various outcomes of literary education, such as types of response to literature, knowledge about literature, attitudes towards literature. These are given in detail on page F-16. These outcomes do not combine to form a single unit of achievement, but they are related to each other in a variety of ways and degrees. An often made comment is that "The more a student knows about literature, the less is his desire to read a book."

HYPOTHESIS II

The various outcomes of literary education are related to, but not identical with each other.

The principal education officers commonly spend much effort on official statements of aims in the various parts of the curriculum. These statements serve a double purpose. On the one hand they explain the purpose of the various subjects to the public, and on the other hand, they seek to influence teaching in the classroom. For literature the latter purpose is often not well served. This study will seek

---

Culture throughout refers not to anthropological culture but to aesthetic culture.

evidence on this particular point and may enable countries to learn from each other what are the most effective ways of setting out objectives.

HYPOTHESIS III

International differences in the stated aims of literary education are related to the place of literature in a nation's culture and practices of literary education.

HYPOTHESIS III-A

International differences in the outcomes of literary education are more closely related to the practices of literary education than to the official statement of its aims.

In the mathematics study it was found that influences outside school were much stronger than had been anticipated. In the case of literature teaching the outside influences, particularly the cultural aspects, might be expected to be highly important.

HYPOTHESIS IV

Within and between nations, outcomes of literary education will vary in relation to social, cultural and individual characteristics of students.

Finally, education in literature is related to the differences or the similarities between an elite literary culture and a popular literary culture. There may be a highly ambiguous and uneasy relationship between these two cultures (e.g., Macbeth and the comic newspaper) and this tension influences the aims, practices and outcomes of literary education, particularly the attitudes of the teacher and the student towards both the elite and the popular culture. While we do not think any simple solution of the problem is to hand, we believe the provision of evidence over a wide range of situations may help towards an ultimate solution.

HYPOTHESIS V

Internationally, the difference between elite and popular culture as represented in the schools is related to the place of each culture in the society as a whole.

The IEA literature committee has tentatively developed instruments to measure the various aspects of the hypotheses. Some of these instruments exist in draft form; some have been specified and we think are capable of being developed; all of them we would like to be subject to the constant scrutiny of the national committees. Drafts will be sent to the committees during the months ahead for comments, suggestions for exclusion, revision and addition. Some of them are:

a. A scheme for content analysis of written responses. This has been substantially developed.

b. Literary preference measures.

c. Literary attitude inventory.

d. Measures of knowledge and application of literary terminology.

e. Measures of preferred responses to literature.

f. Measures of knowledge of allusion to national and cross-national cultural background (e.g., myth and legend).

g. Measures of familiarity with major literary texts, both national and cross-national.

h. Measures of reading habits.

In addition to the stated hypotheses, the IEA literature committee has developed many tentative sub-hypotheses and operating hypotheses. These will be the subject of consultation with the national committees during the next few months.

LITERATURE

INSTRUCTIONS TO NATIONAL CENTERS

We would like to obtain from the National Centers cooperation in the following ways:

1. The appointment of a National Committee in Literature which would include persons who have, among other areas of competence: knowledge about the literary criticism and theory of the nation; familiarity with teacher training practices; familiarity with the various school practices in literature; and familiarity with social science research methods.

2. A review by the National Center of the document entitled "The Goals of Literary Teaching" (attached to this document as Appendix which sets out a summary of educational objectives in literature as to their clarity and accuracy, including suggestions for revision, additions and an indication of the relative importance of the objectives and including an indication on the enclosed grid the national emphasis of various cells. Queries about the grid are expected; we anticipate that it will take three weeks for queries to be made and answered.

3. A criticism by the Center of the major hypotheses of the study as presented in the "Rationale" document, and suggestions as to additional major hypotheses.

4. The submission of copies of official guides for instruction in literature (and/or other official documents pertaining to education in literature), copies of national or regional examinations in literature and any explanatory materials for these exams, titles of the one or two most widely used textbooks in literature (if such exist), and titles of major books on pedagogy in literature. (NOTE: the IEA committee can read documents in English, French or German).

LITERATURE

THE GOALS OF LITERARY TEACHING

The study seeks to describe and to measure achievement in the study of literature. Such description and measurement will be concerned with the following educational goals and sub-goals:

I. Knowledge about literature

a. knowledge of specific literary texts
b. knowledge and application of the terminology of literary study
c. knowledge of the cultural heritage (mythology and prime sources of literary allusion)
d. knowledge of literary history (This sub-goal will not be investigated to any great extent in this study because of extreme national and sub-national differences in the particulars of its literary history and the emphases which literary history receives. National centers are invited to make appropriate case studies with regard to this goal.)

II. Development of Response to Literary Works

Response is defined as the result of interaction between the individual and the literary work he reads. Such results, of course, may be cognitive or affective or both. One of its exponents will be desire to continue reading that work. Response is defined variously as appreciation, taste, or enjoyment. Although it is indescribable in a direct sense, it can be described indirectly through the observation of the works preferred by the individual. The educational goal is generally the development of the threshold of response to include more difficult and "better" literary works.

III. Development of a Pattern of Expressed Response

When an individual reports on his response to select parts of that response which he thinks important to communicate. The expressed response exists in a pattern which can be described, and, in respect to that description measured as to achievement. The means of describing the pattern of expressed response is contained in "The Elements of Writing about Literature."* A major educational goal is the development of a consistent pattern or series of patterns of expressed response (e.g., that the student will analyze works in a consistent fashion, that he will evaluate literary works in a particular way, or that he will view each work in its historical context).

---

*This document may be obtained directly from: Dr. Alan C. Purves, Department of English, University of Illinois, Urbana, Illinois 61801

IV. The Internalization of Positive Attitudes Towards Literature as a Whole and the Application of These Attitudes with Regard to the Work in General

a. The development of the habit of reading

b. The development of specific positive attitudes towards literature and the writer

c. The transfer of expressed response patterns to non-literary events

These goals may be expressed in terms of the grid on the next page. We would appreciate your indicating the emphasis in your curriculum on the cells of the grid in the following manner:

0 no emphasis
1 a little emphasis
2 minor emphasis
3 major emphasis

Content	Be familiar with	Apply knowledge of specific literary texts to	Apply literary history to	Apply literary theory to	Apply cultural information to	Respond to	Develop a consistent pattern of preferences	Express a response to	Express a consistent pattern of responses to	Internalize	Transfer value systems
Specific literary texts											
Literary history											
Literary theory											
Cultural information											

Notes to this grid are given on the following page.

NOTES

Content

Specific literary texts refers to both whole works and extracts

Literary history refers to the facts of authorship, text, intellectual background, and other such information. Behavior with respect to this aspect of content are less important to this study than behavior with respect to other parts of the content for the reasons stated earlier.

Literary theory refers to the terminology and systems of literary criticism and theory.

Cultural information refers to the corpus of mythologic, epic, and folk literature of a culture. Parts of this corpus become a symbol system in the literature of the culture.

Behavior

Respond to refers to the ability to respond to literature and is a major goal, but not one that is measurable except as shown through preference, expression, and consistent pattern of expression.

Develop a consistent pattern of preferences refers to the development of a consistent pattern of choice of one type of literary work over another.

Express a response to may take a variety of forms - verbal and non-verbal.

Express a consistent response to refers to the habit of expressing a preferred type of response to a large number of works.

Internalize and Transfer value systems are applicable to literature as as a whole including texts, history, theory, and cultural information.

LITERATURE CROSSOVER STUDY

1. Introduction

Now that the I.E.A. group is engaged in an investigation of several new subject areas (Science, Reading Comprehension, Literature, Civics, French and English as Foreign Languages), many new problems have been encountered. Within each subject area an international committee of experts has been meeting for the last two years to devise a battery of international achievement tests in their particular field. This work is now nearing completion, although in some of the subject areas many questions are as yet unanswered.

One of the international committees, under the chairmanship of Professor Arthur W. Foshay of Columbia University, New York, has been concerned with constructing instruments to measure the understanding and appreciation of Literature. One important conclusion that they have reached is that there are no specific literary works which would form universal content area. Neither Shakespeare nor Dante, nor indeed any other writer, can be used as a basis for test questions in every I.E.A. participating country. The choice then seems to lie between allowing each participating country to devise its own set of tests, or to introduce examples of literary form which are new to all the students in the study. A common set of questions on these new pieces of literature (common in the sense that they will be the same in each country) can then be put to the students. The literary forms that were considered were poems or brief sections of prose of about 1,000 words in length. Much discussion then centered around whether to use multiple-choice or open-ended test items. The point at issue can be illustrated by an example:

a) In lines 1 to 22 John Deweck has many feelings about what he sees at the fountain. Which of the following best summarizes his feelings?

   (A) He feels that he is separated from other people and that he has lost his youth.
   (B) He feels that his heroic past has been forgotten by others.
   (C) He feels that the park is unattractive when there are people in it.
   (D) He senses that he can be understood better by people his own age.

b) How does he feel about what he sees at the fountain (lines 1 to 22)?

____

____

____

The first form of the question can be machine-scored objectively, reliably and cheaply. The second form requires the subjective judgment of one or more human graders and tends to be more expensive and less reliable. However, the issue of validity was raised by the Committee. Some of them believed very strongly that only the second form of the item could hope to measure the type of instrument in which they were interested. Despite the added inconvenience and expense involved, they felt that a test consisting of open-ended items was the only one that could be of real value. Of course the crucial consideration here was the definition of the achievement that the Committee wished to measure. This was difficult to specify. However, it was found to be possible to agree on the relevance of a certain type of open-ended question. It was proposed by some members of the Committee that carefully designed multiple-choice questions might measure the same dimension and that therefore these multiple-choice questions might replace those of the open-ended variety in the final instrument. To test this proposal, and also to investigate some of the problems in putting together an international test of literary understanding, two small studies were carried out. These are reported below:

2. Poetry Study

a) Design. Two short poems were chosen. These were "Two At Norfolk" by Wallace Stevens, and "Ample Make This Bed" by Emily Dickinson. Each poem deals with the subject of death. Two tests, one in multiple-choice format and one with open-ended questions, were constructed for each poem. Each test required 45 minutes to one hour to complete. A sample of students was drawn and each took two of the tests. The first test was always on the poem "Two At Norfolk," and the second on the poem "Ample Make This Bed." Four pairings of tests were possible for this, and they defined the four treatment groups. The data were subsequently analysed for internal consistency within the tests and correlations between tests on the two different poems.

b) Sample. The sample for this study consisted of 275 students aged 17 or 18. All were drawn from three English grammar schools and may be considered well above average in general ability. However, since they contained substantial numbers of Science specialists, as well as specialists in Literature, it was expected that the sample would show wide score variations on a test of Literature. Both poems were by American authors, and therefore presumably unfamiliar to most of the sample. The students were randomly divided into four treatment groups and each took two tests as outlined above.

TABLE 1: SAMPLE SIZE AND INTER-POEM CORRELATIONS

	RESPONSE MODE			
	Ample Make This Bed	Two At Norfolk	N	r
Group 1	Open-ended	Open-ended	63	.245
Group 2	Multiple-choice	Multiple-choice	70	.212
Group 3	Open-ended	Multiple-choice	73	.193
Group 4	Multiple-choice	Open-ended	69	.275

c) Results. The sample size and correlations between the two poems are shown in Table 1, while Table 2 gives more general statistics for the four tests. The results for the open-ended tests are the averaged grades assigned by two independent markers. The grades for these markers related with each other to the extent of 0.96, a surprisingly high figure which suggests that a very adequate scoring scheme had been worked out beforehand. As the markers did not produce sub-scores for different parts of the test, it was not possible for these tests to estimate their internal consistency.

TABLE 2: TEST STATISTICS FOR POETRY

RESPONSE MODE	POEM	N	Possible Score	Mean Score	S.D.
Open-ended	Ample Make This Bed	136	29	6.83	4.5
Open-ended	Two At Norfolk	132	31	8.47	4.0
Multiple-choice	Ample Make This Bed	139	12	8.22	1.65
Multiple-choice	Two At Norfolk	143	15	9.26	2.45

For the multiple-choice tests it was possible to estimate reliability using KR 20. The estimates obtained were, for "Two At Norfolk" R =0.54; for "Ample Make This Bed" R = 0.28.

In general these results were rather disappointing. Although the data provided no evidence to suggest that open-ended and multiple-choice tests were measuring fundamental different aspects of ability, the uniformly low values of the correlations between the poems (whatever the form of the tests) seemed to suggest that the rankings obtained for a group of students would be very dependent upon which poem was chosen for the test. Clearly one would not, in these circumstances, wish to allow each individual country within the I.E.A. project to nominate its own poem. To some extent these low correlations may be produced by the rather low values of the test reliabilities. In fact, if we assume that the open-ended tests had internal consistencies similar to those for the multiple-choice tests, then low correlations between poems would be very largely explained. However, in these circumstances, it might be thought that none of the tests used provided an adequate measure of literary understanding.

3. Prose Study

a) Design. A cross-over design similar to that employed in the poetry study was used. Two short stories, "The Man by the Fountain" by Georges Hebbelinck and "My Childhood" by Maxim Gorki, formed the basis of both multiple-choice and open-ended tests. Each story was approximately 1,000 words in length. The tests were labelled 7, 8, 9 and 10. Each required up to one hour to complete. Students were asked to take two of the tests, one on each of the two short stories. Once again the stories were expected to be unfamiliar to the students. Four treatment groups were defined as in the previous study.

b) Sample. Four separate samples were drawn:

(i) 376 English school children drawn from one grammar and three schools. All were in the fourth year of secondary school with an average age of about 14 years 6 months.

(ii) 289 English students drawn from the Sixth Form of a grammar school and a Technical College. The mean age is 18 years.

(iii) 338 American High School students selected from the ninth grade, and aged about 14.

(iv) 289 students from the American twelfth grade aged about 18.

(The tests had been designed and constructed for use with two populations of students, one aged 14 and the other 18). As it was soon apparent that there were considerable differences of performance between samples, no pooling of data took place and the results for each sample are reported separately. Each sample was divided randomly between the four treatment groups. The actual numbers assigned to each treatment are given in Table 3.

TABLE 3: RESPONSE MODE AND SAMPLE SIZE BY COUNTRY AND AGE

	RESPONSE MODE		ENGLAND & WALES		U.S.A.	
	The Man by the Fountain	My Childhood	14 yr.	18 yr.	14 yr.	18 yr.
Group I (7,9)	open-ended	open-ended	92	66	80	66
Group II (8,10)	multiple-choice	multiple-choice	86	78	83	80
Group III (7,10)	open-ended	multiple-choice	90	85	86	62
Group IV (8,9)	multiple-choice	open-ended	108	60	89	81

c) *Results.* The simple test statistics are presented in Table 4. It is interesting to note the higher average scores and lower standard deviations of the American students on each of the four tests. It will, of course, be necessary to wait for the full scale I.E.A. study to determine if this finding holds for truly representative samples.

TABLE 4: TEST STATISTICS FOR SHORT STORIES

TEST	RESPONSE MODE	SHORT STORY	SAMPLE	N	MEAN SCORE	S.D.
7	open-ended	The Man by the Fountain	England 14 yr.	182	9.84	5.18
			England 18 yr.	151	14.80	7.46
			U.S.A. 14 yr.	166	14.36	3.99
			U.S.A. 18 yr.	128	18.30	5.77
8	multiple-choice	My Childhood	England 14 yr.	194	12.45	4.59
			England 18 yr.	138	15.91	4.45
			U.S.A. 14 yr.	172	17.40	3.26
			U.S.A. 18 yr.	161	18.37	2.88
9	open-ended	The Man by the Fountain	England 14 yr.	200	11.92	6.18
			England 18 yr.	126	13.94	7.82
			U.S.A. 14 yr.	169	16.40	3.53
			U.S.A. 18 yr.	147	18.85	4.74
10	multiple-choice	My Childhood	England 14 yr.	176	11.93	4.86
			England 18 yr.	163	14.38	5.78
			U.S.A. 14 yr.	169	16.40	3.53
			U.S.A. 18 yr.	142	15.59	3.67

(The Maximum score possible on the open-ended tests is 35, and on the multiple-choice tests, 23)

Although there are striking similarities in the score distributions (mean and standard deviation) of the four tests, this result must be regarded as entirely fortuitous. The maximum possible score for the open-ended tests is 35 as against 23 for the multiple-choice form.

Each of the test booklets 7 and 9 was graded by three markers. The three who completed most of the work correlated 0.92, 0.93 and 0.93 with each other, which, although not as high as the value achieved in the poetry study, must be regarded as very satisfactory. One marker had to drop out before the end and was replaced for the last 25% of the scripts. The substitute marker correlated 0.88 and 0.90 with the remaining two.

On this occasion the markers produced sub-scores for different sections of the test and it was thus possible to estimate reliability through internal consistency for all four tests. These estimates are given in Table 5. They appear to be reasonably consistent, although there is some suggestion that reliabilities were higher in England and Wales, and this is no doubt due to the greater variability of these samples.

TABLE 5: RELIABILITY ESTIMATES (KR 20)

Test	Short Story	Response Mode	England & Wales 14 yr.	England & Wales 18 yr.	U.S.A. 14 yr.	U.S.A. 18 yr.
7	The Man by the Fountain	open-ended	.69	.85	.52	.75
8	The Man by the Fountain	multiple-choice	.81	.82	.68	.65
9	My Childhood	open-ended	.79	.83	.63	.62
10	My Childhood	multiple-choice	.83	.90	.74	.78

TABLE 6: INTER-TEXT CORRELATIONS

Group	Response Mode		England & Wales		U.S.A.	
	The Man by the Fountain	My Childhood	14 yr.	18 yr.	14 yr.	18 yr.
Group I (7,9)	open-ended	open-ended	.60	.43	.35	.41
Group II (8,10)	multiple-choice	multiple-choice	.64	.66	.45	.75
Group III (7,10)	open-ended	multiple-choice	.48	.50	.58	.60
Group IV (8,9)	multiple-choice	open-ended	.46	.51	.39	.49

The correlations between the pairs of texts taken by different students are presented in Table 6. There is a clear contrast with the values obtained earlier with the poetry texts. Although in some cases these correlations are not very high, there is no evidence to suggest that the size of the correlation depends upon the style of response to either short story. "Open-ended" correlates with "open-ended" as well as "multiple-choice" with "multiple-choice," or indeed with a mixture. In Table 7 these inter-text correlations have been attenuated to allow for unreliability in the two measures. The resulting figures are estimates of correlations between the tests under conditions such that each test had perfect reliability. These correlations are, of course, somewhat higher than in Table 6, but it is interesting that no clear pattern emerges. These correlations average about 0.70 - which suggests that after variance due to unreliability is removed, roughly half the remainder is shared between the tests and half is specific to individual tests.

TABLE 7: INTER-TEXT CORRELATIONS ATTENUATED FOR UNRELIABILITY IN THE TWO MEASURES

Group	Response Mode		England & Wales		U.S.A.	
	The Man by the Fountain	My Childhood	14 yr.	18 yr.	14 yr.	18 yr.
Group I (7,9)	open-ended	open-ended	.81	.52	.62	.60
Group II (8,10)	multiple-choice	multiple-choice	.78	.70	.63	1.00
Group III (7,10)	open-ended	multiple-choice	.62	.58	.91	.79
Group IV (8,9)	multiple-choice	open-ended	.58	.62	.60	.77

4. Conclusions and Recommendations

The results presented on the preceding pages offer no evidence that multiple-choice questions per se measure anything different than open-ended questions per se. It will be recalled that this was the single most important point to be tested in these studies. Although the correlations between multiple-choice and open-ended tests are not high, they are not perceptibly lower than are correlations between multiple-choice tests themselves or between open-ended tests.

It is possible, however, to distinguish between multiple-choice and open-ended tests on financial grounds. While the multiple-choice responses were scored by computer after they had been punched on cards, the open-ended responses were submitted to expert markers. The cost of grading one open-ended script (using three markers) turned out to be about 40 cents. The computer cost per script was about one cent, to which must be added approximately 2 cents for key punching. It is clear that the multiple-choice forms of the test will prove much cheaper if large numbers of students are involved. Although in view of the high correlations between the markers, relatively little reliability would have been lost by using only one marker instead of three, it would not appear possible to reduce the cost per script to much below 13 cents, which is four times the cost of scoring a script of the multiple-choice format. The high inter-marker correlation was no doubt due to the very careful briefing that the markers were given before they commenced work. They had an opportunity to examine a selection of scripts and to discuss the marking procedures in very great detail. This is rather time-consuming, of course, but it does demonstrate that with careful preparation a high inter-marker correlation could be achieved. With only one marker there would, of course, be no check on the reliability of the grades produced. The results of these two studies suggest strongly that only multiple-choice forms be used for large-scale testing in this field.

In each of the two studies here reported, it would seem that at least half the non-error variance can be attributed to characteristics specific to the text under consideration. Less than half the total variance can be attributed to a general measure of ability in the literary field. This will be an obstacle to the construction of instruments designed to provide highly reliable scores of general validity. While it would be possible to construct a multiple-choice test to measure reliably the understanding of a specific short story or poem (by including a sufficiently large number of questions), a reliable measure of general ability to understand and interpret literature could only be based on responses to a wide selection of literary works. It may well prove to be a problem here, since the time taken to complete the test depends in large part upon the amount of textual material which the student has to master. The finding that we get reliabilities of about .75 from tests one hour long is frankly discouraging.

In view of the low correlations between texts, it would seem very unwise to try to compare scores made on different texts in different countries. If the I.E.A. project is to proceed with measures in the field of literary understanding, it seems clear that the texts used must be the same in all the countries involved.

APPENDIX G

FRENCH DOCUMENTS

APPENDIX G1

FRENCH AS A FOREIGN LANGUAGE RATIONALE

This document is intended to indicate to the National Centers and others who may be interested, the overall rationale for attempting to study achievement in French as a foreign language on an international basis. It consists of four parts:

I. Design of the study (from a statistical point of view)
II. Hypotheses
III. Variables
IV. Target Populations

The section on hypotheses and variables will not include a discussion of criterion variables, since this will come out of the document on content analysis intended for the National Centers. It may be assumed, however, that the study will eventually involve the development of an extensive battery of tests measuring various aspects of achievement in French as a second language.

I. Design of Study

1. The main interest of the study will lie in the results which will be obtained with comprehensive tests of attainment in French. These we hope to administer to representative samples of three levels of the populations of French learners: The pre-university, the fourteen year-olds, and the ten year old levels (for further details of these samples/see below). Since the same set of French tests will be administered in all participant countries, and since we hope to have statistics about the proportions of students studying French at each level in each country, we hope as a result of the study to be able to estimate for each country the total number of students who know French well enough to satisfy certain standards defined by performance on the tests. Thus we hope the study will reveal the "yield" of French teaching in each country.

2. Apart from determining the "yield" of French teaching, however, the study has many other aims mostly relating to factors which affect the extent to which French teaching is successful. Several considerations have led us to the conclusion that these aims are unlikely to be achieved by comparing the overall means, or even the means of selected subgroups, across countries. In each country French is an optional subject, so it is unlikely that we shall be able to find out whether students who take French are representative of the school-going population as a whole. For this reason alone it would be quite unsafe to assume that the populations of French learners in the participant countries are comparable in general ability or in ability to learn French. Furthermore, countries vary in cultural background, in native language, and in the orthography they employ with the result that the task of learning French varies in difficulty from one country to the next.

3. Bearing all this in mind we propose to study in each country the extent to which attainment in French is dependent on a large number of interesting and important variables, and then go on to compare the influence of these variables across countries. These variables fall under seven heads:

a. Pupil variables (e.g. sex, number of hours spent in French class)

b. Parent variables (e.g. interest in and knowledge of French)

c. Teacher variables (e.g. extent of professional training in French and in methods of teaching French)

d. Method and Course content variables (e.g. stage at which each of the four major areas of French learning--listening, speaking, reading and writing--is introduced)

e. School variables (e.g. Streaming, size of class)

Besides these variations within countries, there are variations amongst countries which must be studied for the purpose of comparing across countries the effect of the variables already mentioned.

f. Mother tongue variables (e.g. the extent to which the orthography, phonological, syntactic and lexical systems differ from French)

g. Country variables (e.g. proximity to or remoteness from a French-speaking community)

It is reasonable to hypothesize that each of these variables contributes to some extent to the score which children obtain in French tests, though of course, the study will determine whether in fact they do contribute and if so, to what extent.

4. In a study comprising so large a number of variables the most convenient way to tease information from the data is regression analysis. This will involve calculating for each country separately the regression of French scores on the "predictive" variables (i.e. variables "a" to "g".) The resulting regression coefficients will indicate the accuracy with which French scores can be predicted, granted that one knew the predictor measures. Furthermore, each coefficient will indicate the predictive power of a particular predictor measure independent of all the other predictor measures. This in effect is the equivalent of its predictive power when all the other variables are held constant.

Once the regression analysis for each country has been completed it will be possible to compare regressions across countries. If regressions are constant, the study will have yielded information of great importance in coming to a further decision: whether the relationship between a predictor variable and French scores is a causal one or not.

If the regressor is constant across all countries, despite considerable variation in the settings in which children study French, and if there are good a priori reasons for considering the relation as causal, then the case for a causal connection is greatly strengthened. In this event the study will indicate which steps aimed at improving the study of French are likely to be rewarded. (Of course underlying certain variables - e.g. sex of student - are factors which cannot be altered. These variables provide information which will be of interest to educationalists and psychologists.) On the other hand, if regression varies from one country to the next, and in particular if the coefficient associated with some predictor is zero in some countries, the connection between the factor underlying this variable and attainment in French is likely to be a complex one. This in turn would mean that measures to improve French teaching directed towards this factor may not have the desired result, or may only produce it when other factors interact with the changes in some unknown manner. In short, it would be wise to delay such measure until further research had determined under what circumstances the factor had a bearing on instruction in French.

II. Hypotheses

The following are examples of hypotheses which might be tested:

1. (i) Variation in achievement profile will reflect variation in teaching methods and objectives: children taught French by audio-visual methods will reach a higher level of achievement in speaking and listening skills, children taught French by grammar-translation methods will reach a higher level of achievement in reading and writing skills.

(ii) This variation will not persist into the pre-university grade (if length of exposure to French is held constant.)

2. Level of Achievement in French will be related to length of exposure to French rather than to chronological age or grade level.

3. Attitudinal factors (in both home and school environment) will affect level of achievement in French.

4. Socio-economic factors (parental education, socio-economic skills, etc.) will be related to child's level of achievement in French.

5. Girls will reach a higher level of achievement in French than will boys: this difference will be accentuated in single-sex schools.

6. Size and composition of French class will be related to achievement in French.

7. Selection procedures will be related to level of achievement in French: children in streamed schools will achieve a higher mean score than will those in comprehensive systems; the latter will produce a higher total "yield" but a lower average score.

8. Children's level of achievement in French will be related to the teacher's training, qualifications, and teaching experience.

9. Variation in economic support for education will be related to variation in level of achievement in French.

10. Children's level of achievement in French will be related to parental aspirations and attitudes.

11. Level of achievement in French will be related to opportunities for contact with French speaking groups.

12. Level of achievement in French will be related to the "linguistic distance" between French and the mother-tongue.

13. Measures of verbal ability in the child's mother-tongue will correlate highly with measures of achievement in French at the advanced levels; in the early stages of learning, there will be a much lower correlation.

III. Variables

The following are a very preliminary listing of some of the variables likely to be used in hypotheses testing.

A. National Variables

1. National Educational System
    - a. Types of schooling available.
    - b. Percentage of each age-group in school
    - c. Percentage of each age-group in the different types of school
    - d. Age of beginning school - Minimum school-leaving age
    - e. Expenditure on education and educational research
    - f. Perceived status of teaching profession
    - g. Type of administration: central/local
    - h. Inspection system: type and extent
2. Specific linguistic factors
    - a. Number of native languages spoken in the country
    - b. History and extent of French teaching in the country

c. Social factors influencing attitudes towards French: religious, economic, etc.

d. Official policy towards the teaching of French: stated objectives

e. Status of French as an occupational or examination requirement

f. Estimate of "linguistic distance" between French and the mother tongue and identification of major sources of interference: reading direction, orthography, etc.

B. School Variables

1. Type (primary, secondary - state/independent, etc.)
2. Sex (boys', co-ed, girls')
3. Size: total enrollment; size of classes; number of classes; age-range in each class; number of teaching staff
4. Method of selection or recruitment
5. Organizational pattern (streamed/comprehensive)
6. Location: urban/rural
7. Curriculum: sugjects taught; choice possibilities degree of specialization; number of languages taught; stage at which languages introduced and to whom (whole ability range?) status of French in curriculum; French used as a medium of instruction for other subjects?
8. Contacts with France and the French: school visits; "twinning"; club activities; availability of French books, films, etc.

C. Teacher Variables

1. Personal data

a. Age

b. Sex

c. Professional training: length and type; formal qualifications

d. Extent and type of training for French teaching; residence in France? Is training perceived as relevant to French teaching objectives?

e. Number of classes taught; size of classes; age-groups within classes; selection of children within classes; time allocated to French lessons. If teacher in primary school, class teacher or French specialist? If class teacher, how many hours is teacher absent from own class to teach French? Is class taught as a whole or in groups?

f. Support for French teacher: in-service training; access to language laboratory; contact with French-speaking groups; availability of French books, films, etc.

g. Teacher's salary scale

h. Teacher's attitudes towards France and the French. Perceived objectives of teaching French

2. Methods and Materials

a. Methods used to teach French: direct/indirect/combination of methods, etc. (Possibility of developing a scale to specify teaching methods with reasonable accuracy.)

b. Sequence of French teaching: listening, speaking, reading, writing? Investigation of specific teaching emphases

c. Course material used: audio-visual course, textbooks, teacher's own materials, etc.

d. Availability and use of funds and equipment (language laboratory, tape recorders, film-strip projectors, etc.)

D. Pupil Variables

1. Age

2. Sex

3. Socio-economic background: parental occupation; level of parental education; size of family, etc.

4. Linguistic background: use of French in the home; contact with French outside school program; number of languages spoken by the child; number of languages learnt by child in school

5. Age at which child introduced to French; number of hours' exposure to French teaching (split, if possible, into the amounts of time devoted to each of the four basic skills)

6. Motivation of child for learning French: perceived objectives for study of French; level of aspiration; attitudes towards French language and people; attitudes towards child's own culture

7. Child' perception of parental attitudes towards education in general and towards learning French in particular (investigation of parental attitudes desirable, if feasible parents' interest in and knowledge of French; attitudes towards child's learning French; perceived objectives for study of French)

8. Child's verbal ability in mother tongue

9. Child's disabilities of hearing, vision, or speech

10. Child's homework; total amount and amount of French homework

IV. Definition of the Target Populations

The design of the study requires that representative samples of students of French be drawn from each of the participating countries in such a way that there will be a maximum degree of comparability among the samples, within and among countries. This "maximum" degree of comparability may in fact be quite limited because of the varying policies concerning French instruction in the different countries or in different sectors of the educational systems of those countries. Nevertheless it will be important and desirable to attain whatever degree of comparability is possible.

The experience of the IEA study of mathematics suggests that the chief way in which some degree of comparability can be attained is to sample with respect to the age or grade of the student. But it is not practicable or even desirable to sample at all ages. Only two or three age-level cohorts can be sampled.

A major constraint upon the sampling which is particularly relevant in the case of an "optional" subject like French is the amount of time the student has been exposed to the language. It is pointless to try to measure achievement in French for a student who has not taken any instruction in the subject. A useful measurement for the purposes of this study can be taken only after the student has had approximately two years, or about 160-180 contact hours of instruction. (For all students included in the study, information will be gathered on the amount of instruction, in years and/or contact hours.) The samples would also include, wherever possible, students who have continued the study of French for more than two years. For example, a sample of students at a given age-level, say 16 years, would include some who would have studied French for as long as 5 or 6 years. (It would not include, however, students who started 5 or 6 years earlier, studied for 2 years, and then dropped it.)

To enable the International Committee to determine appropriate age-level samples, therefore, country-by-country statistics are needed on the numbers of students of each age-level who have studied French for at least two years and who are still enrolled in French. If it is impossible or difficult to collect statistics by age-level, statistics according to grade-level are acceptable, even if only in terms of percentage.

An example of the sort of table that is desired from each participating country is given on the next page.

Age (or grade level)	Total number of pupils in the age-level (or grade level) whether studying French or not	(Estimated) Number (or Percentage) of students who enrolled in French and who have studied for:					
		2 yrs	3 yrs	4 yrs	5 yrs	6 yrs	. . .
8 ( )							
9 ( )							
10 ( )							
11 ( )							
12 ( )							
13 ( )							
14 ( )							
15 ( )							
16 ( )							
17 ( )							
18 ( )							

It may be useful to set forth a tentative set of target populations based on the Committee's general impressions of instruction programs in the several participating countries. Three sampling points seem to be suggested:

Age 10 (approximate grade level: end of grade IV) Because of the Committee's interest in the teaching of French in the elementary school (as a result of the conclusion of the May, 1966 conference at the Unesco Institute for Education), a sampling at this age level would yield information concerning how much can be achieved at this age level and what factors are associated with it. It might not be possible to collect samples of this age level in all participating countries, however.

Age 14 (approximate grade level: end of grade VIII) The participating countries have programs which introduce French as a foreign language anywhere from grade IV to grade VII. By the end of grade VIII there would be in each country sizable numbers of children who have studied French for at least two years.

Age 17-19 (pre-university level) In each of the participating countries, there would be substantial numbers of students who would have studied French as a foreign language for at least two years and are still enrolled in it. Of those enrolled, some would have studied for many more than two years.

APPENDIX G-2

French Test Rationale and Description of Pre-Tests

General Considerations

The test descriptions presented here reflect in large part the tentative framework for test development previously described in the minutes of the I.E.A. International Sub-Committee meeting held at Educational Testing Service, Princeton, New Jersey, from July 31-August 4, 1967. A meeting of the U.S.A. National Committee for French was held in Princeton some weeks later (September 18-19, 1967), and this committee reviewed the proposed specifications in considerable detail. On the basis of these meetings, and taking account of various other comments and suggestions sent in by other national committees, Mrs. Clare Burstall and Dr. John Clark assembled several hundred draft test items for presentation at the meeting of the full International Committee held in London from October 9-13, 1967. During this meeting, considerable discussion was raised concerning the general theory and orientation of the tests, and certain modifications were suggested for some parts of the tests. A detailed description of the various points raised is presented in the minutes of the International Committee meeting, and will not be repeated here. In addition to the general discussions, the Committee also edited the actual test items and checked the stylistic and technical accuracy of these materials.

The test descriptions and rationale presented below are intended to show the composition and measurement functions of the pre-test materials as they have been conceived and assembled to date. These materials are to some extent a synthesis of two basic approaches to foreign language testing: on the one hand are included various item types which attempt to measure the student's general command of the language as reflected in his ability to comprehend spoken or written materials presenting various "real-life" linguistic situations, as, for example, conversations between two or more people, radio broadcasts or loudspeaker announcements, and representative reading texts of paragraph length or longer. On the other hand, an attempt has also been made to provide a number of items which lend themselves to close analysis in terms of the linguistic features which they incorporate. Items of this type can serve a diagnostic function in that they allow the tester to determine specific language behaviour which students have acquired or have failed to acquire in their course of study. Items testing the accurate pronunciation of certain French phonemes or the written mastery of certain verb forms would exemplify this second approach. It seems fair to say that both approaches to evaluating language competence are valid in their own right and serve a complementary function in the overall testing endeavour.

A number of potentially useful testing ideas which are not reflected in the draft tests have been presented, both orally during the various meetings and through correspondence with the national centers. We wish to acknowledge receipt and consideration of these suggestions; however, in view of the restraints imposed on the testing program, both in terms of the total testing time available and also in consideration of the problems involved in the controlled administration of certain types of test materials, it has been necessary to adopt a rather economical approach in selecting item types and content for inclusion in the tests. A number of interesting but somewhat peripheral aspects of language competence (for example, the ability to deal with dictee materials, or to read poetry aloud) have necessarily been omitted in favour of closer attention to the more basic language skills.

Description and Rationale for Individual Tests and Item Types

POPULATION ONE TESTS

The tests for Population One (10-year-old level) will be administered only in England, Scotland, and the U.S.A. At this level, most students will not have had sufficient training in writing to justify the inclusion of a Writing test. Thus, tests for Population One will be limited to Listening Comprehension, Reading, and Speaking (N.B. - For Population One, as well as for each of the other populations, the Speaking tests will be administered only to a sub-sample of students in view of the administrative problems involved in tests of this type).

Listening Comprehension Test: The basic testing technique will be the tape-recorded presentation of single sentences, spoken by a native French speaker using a "standard" style of pronunciation (i.e , free from marked regional characteristics). Speed of delivery will be sufficient to insure natural liaison and intonation. After listening to each sentence, the student will choose from a set of four pictures printed in his test booklet the one which correctly depicts the spoken sentence.

It should be noted that this technique eliminates any element of reading comprehension; this is particularly important of the 10-year-old level since, depending on the training program followed, many students will have had little exposure to reading by this time.

Depending on the ways in which the answer options are planned and manipulated with respect to the spoken material, the four-picture format can be used to test both vocabulary knowledge and syntactical control, particularly for number, gender, and subject-object distinctions.

The Listening Comprehension test will require about 30 minutes, including administration time, and will contain 50-60 items. At this level, students will mark their answers directly in the test booklet rather than use a separate IBM answer sheet. A short test break is anticipated about half way through the test.

Reading Comprehension: The Reading test at this level will also make use of a 4-picture format, except that the stimulus sentences will now be printed in the test booklet in boldface print. In the final section of the test, a few brief passages of continuous prose will be presented, and each passage will be followed by two or more printed multiple-choice questions based on the passage. This section is intended primarily for the most advanced students in the sample, and its inclusion will help to avoid the possibility of a "ceiling effect" for the test.

The lexicon used in both the Listening and Reading tests has been drawn from Francais Fondamental, ler degre. An attempt has been made throughout to use only those vocabulary items which could be presumed common to the various teaching programs used at this level.

The Reading test will require about 20-25 minutes overall and will contain approximately 50 items.

Speaking: The Population One Speaking test will consist of three sections, Pronunciation, Structural Control, and Fluency. In the Pronunciation section, the student will listen to and initiate a number of brief but complete and meaningful sentences (e.g., "La neige tombe"). The sentences are selected to exemplify selected details of pronunciation (principally, the accurate production of vowel and consonant phonemes) and intonation (interrogative, declarative and exclamatory patterns). Although it might be possible to score the production of each sound or intonation pattern along an extensive continuum of quality, it is considered more practical and reliable to use a simple three-point scale, or perhaps, simply to score the items as right or wrong.

The Structural Control section presents a number of single pictures depicting a simple scene (e.g., a child drawing a picture). The student is asked to reply orally simple questions relating to each picture (e.g., Que fait la petite fille?). The pictures and questions are designed so that only one specific answer or a very restricted range of answers would be elicited. The emphasis of this section is thus not on the student's ability to "talk" in French from a fluency point of view but rather to produce an appropriate and grammatically correct response to a well-defined stimulus.

The final section of the test - Fluency - presents single but fairly complex pictures which the student is asked to describe in some detail. Both overall fluency and grammatical accuracy will be scored according to 3- or 4-point scales defined in verbal terms.

The Speaking test will be administered on an individual basis using two tape recorders, one for presentation of the instructions and aural stimuli and one to record the student's responses. Centrally located scoring, possibly at Education Testing Service, is envisioned for the Speaking tests at all levels.

The timing for the Population 1 Speaking test will be 10-15 minutes, including instructions.

TESTS FOR POPULATIONS 2, 3, and 4

At the present time, draft tests have been prepared in Listening, Speaking, Reading, and Writing at the Population 2-3 level (both of these populations will take the same tests) and at the Population 4 Advanced ("specialist") level. General survey tests at the Population 4 level have not yet been assembled, since test development has proceeded on the assumption that the survey test will be made up, for the most part, of selected items from the Population 2-3 and Population 4 advanced tests.

Within a given skill area, the same general item types have been used for both the Population 2-3 and Population 4 Advanced tests, although the general level of item difficulty is of course greater in the latter case. The format descriptions which follow may be considered as applying to the tests at both levels.

Listening Comprehension: The first section (LC/PICT) of the test presents 4-picture panels together with an aural stimulus of sentence-length. Basic vocabulary is tested, as well as certain important syntactical elements (see, for example, items 1, 4, 107, 108). These picture items are intended to provide a pleasant and interesting introduction to the Listening test as a whole.

The second section (LC/SENT) consists of a number of simple statements or questions such as would be encountered in everyday conversational situations. In the student's test book are printed four short statements, one of which represents the most normal or appropriate reply to the spoken sentence.

It is realized that this technique involves a reading comprehension element, which would not be the case if the answer options were themselves spoken (as had originally been planned in the July 31-August 4 sub-committee report). However, it became apparent upon discussion that the use of spoken answer options was problematical for several reasons. First, since the spoken options would necessarily have to follow one another on the tape, students might be able inadvertently or deliberately to signal the correct answer either by marking their answer sheets immediately after the correct answer is given or by making some other sort of gesture. Second, the spoken-option technique was considered somewhat uneconomical of testing time, in that a total of 5 sentences would have to be read aloud in order to present a single item. For these reasons, the committee considered that it would be more useful to present the options visually, with the proviso that the answer options themselves be couched in such simple terms that the student for whom the Listening test per se would be appropriate would not have difficulty in reading the answer options. Persons reviewing the listening items are urged to keep this aspect in mind, and to point out any items which they would consider to present an unfair reading load.

The third section of the Listening test (LC/CONVER) is similar to the second except that now the student listens to a short conversation between two people and checks the option which correctly describes what is happening in or is implied by the conversation. Depending both on the linguistic complexity of the conversations themselves and the particular questions asked, the level of difficulty of these items can vary from extremely simple and straightforward to highly challenging. (As a general rule, the items at the beginning of any given section are deliberately quite easy, and gradually increase in difficulty throughout the section, although occasional easier items are interspersed to keep the weaker students from becoming discouraged.)

The fourth section of the test (LC/PASSAGE) present short radio or loudspeaker announcements, reclames, and so forth, followed by a spoken question on the overall situation or on some detail of the situation. Again, the difficulty level of the items is varied within the section.

The fifth and final section (LC/LONG CONVER) consists of longer conversations or dramatic scenes which continue for one or two minutes, and which the student is expected to listen to as a whole. Following this material, the student is asked a number of questions based on the passage. The questions are intended to test basic comprehension of the passage, to the extent that it would be retained by an interested native speaker, as contrasted to the recollection of minute details which even the native speaker would probably not remember.

With few exceptions, the items sppearing in all sections of the listening test should be considered as testing general comprehension at various levels of linguistic complexity, rather than the control of certain specified elements of phonology, syntax, or lexicon.

Reading Tests: The Reading tests, as presently established, consist of two sections: vocabulary-in-context (RC/VOCAB) and reading passages (RC/TEXTS). Although it would appear that the comprehension of written texts, as tested in the second section, is closest to the tasks facing the student in real-life situations, there is a slight drawback in the use of items of this type (or at least, in their exclusive use) in that it is difficult to sample within any reasonable amount of testing time all the levels of difficulty and extent of lexicon which it would be useful to cover. As a consequence, a preliminary section consisting of a fairly large number of discrete reading items, generally one or two sentences in length, has been provided. These so-called "vocabulary-in-context" items, present a short passage or sentence from which a single word or group of words has been omitted. It is the student's task to select from among four printed alternatives the word or phrase which best fits into the stimulus sentence. This item type lends itself to the testing of a number of important reading elements, including the vocabulary of nouns (of various degrees of difficulty), adjectives, adverbs, verb forms, and idiomatic constructions. In some items, the point of difficulty lies in the options themselves, whereas the stimulus phrase itself is quite straightforward; in others, the four options presented are easy to understand, but in order to choose the correct option, a close and accurate understanding of the stimulus phrase is required. Rather than detailing here the various possibilities of this item type, it is suggested that the test reviewer look through these items at leisure to determine the different aspects of comprehension involved.

As previously stated, the second section of the test presents longer reading passages of various degrees of complexity and in varying written styles. Two or more printed questions, each with four answer options, are based on each passage. Since in the Reading test the student is able to refer back to the passage (as contrasted to the situation for listening comprehension passages), it is here possible to ask detailed questions about the content of the passage, as well as to ask more general questions on the theme or mood of the passage.

The Reading tests are anticipated to each require about 30 minutes' testing time, including instructions. As they are here presented, there is a slight surplus of Reading test items; the final forms will probably incorporate somewhat fewer items in each of the two sections.

Writing Tests: Unlike the Listening and Reading tests, which are for the most part addressed to problems of general comprehension, the Writing tests offer the possibility of testing the student's command of specific details of syntax, morphology, and lexicon. It should also be pointed out that the Writing tests measure the student's language proficiency from a *production* point of view (as do the Speaking tests), whereas the Listening and Reading tests constitute measures of *receptive* skills.

In specifying item types for the Writing test, it was considered essential to allow the student to construct (that is, actually write out) his responses, rather than to choose from among multiple options. On the other hand, it was necessary to insure - for the sake of scoring reliability - that student would respond uniformly and predictably to the materials presented (assuming, of course, that they had a correct knowledge of the points being tested). Thus, the first and major section of the Writing tests (WR/GRAM) at both levels was designed to present partially completed sentences, each of which omits a single word (noun, verb, adverb, adjective, conjunction, preposition) bearing on the particular syntactical or morphological problem being presented. Although the student does not in this part of the test produce complete sentences, he is nonetheless required to produce that portion of the sentence which is considered to represent the essential problem. Since the sentences are relatively short and require only one-word fill-ins, a fairly large number of such items can be presented in a reasonably short time span.

The second section of the Writing tests (WR/PRON) is designed to test the student's ability to handle various collocations of subject and object pronouns, both for proper form and correct location in the sentence. The general technique is to present a model sentence containing a number of nouns (e.g., "Est-ce que Jacques prete sa bicyclette a son frere?") and to ask the student to rewrite the sentence substituting the correct pronouns ("Oui, il la lui prete"). The difficulty level of the items is determined by the number of nouns to be replaced, as well as by the form of the sentence (affirmative or negative) and the tense of the main verb (present or passe compose).

The third section of the test (MR/VBS, MODIF) evaluates the student's ability to produce correct verb forms (regular and irregular) for various persons, and to form correctly various descriptive adjectives, both regular and irregular. The testing technique is to present a model sentence, and then to change one or two words (e.g., from "homme" to "femmes"); this change in turn requires changes in verb forms and/or adjectives in order to produce a correct sentence.

The final section of the Writing test (WR/DIR COMP) gives the student the opportunity to produce a "free composition," albeit within certain necessary limitations. The general testing procedure is to give a series of stimulus-words which suggest a short story. The student writes out his story, using the stimulus words presented, and in their order of appearance. This section of the test is scored for inclusion of all the necessary elements, and for the grammatical and stylistic accuracy of the sentences produced. Although this last section is the least "objective" in format, prior experience with this item type indicates that it can be evaluated by properly trained scorers with a satisfactorily high degree of reliability.

Speaking Tests: The Speaking tests at the higher levels correspond in general outline to the Population One Speaking test, except that certain other item types are added. Following the Pronunciation section appears an Oral Reading section in which the student reads a short paragraph aloud. Scoring is based on the accurate pronunciation of certain phonemes and on proper intonation, but particular attention is also given to required liaisons, across-sound modifications (such as the denasalisation of un bon ami), and other such features of the "Chaine parlee."

The Structural Control section is basically similar to that for Population One, except that the pictures and questions are designed to elicit more complex responses.

The Fluency section is broken into two sub-sections: the first subsection presents a panel of three pictures depicting a simple event. The student is asked to "tell the story that the pictures tell," and it is anticipated that the responses will all be in the present tense. In the second sub-section, a single picture is shown and the student is asked to describe what probably happened prior to the time of the picture, what is happening at the present moment, and what the probable outcome will be. This technique allows the student somewhat more response latitude, and at the same time checks his command of past and future tenses.

An effort will be made to hold the higher level tests within a 15 minutes total time limit, including the time required for the additional sections.

APPENDIX H

CIVIC EDUCATION DOCUMENTS

APPENDIX H-1

CONCEPTUAL FRAMEWORK
FOR
CIVIC EDUCATION

As a result of the recent IEA International Civic Education Committee meeting held at the end of April, 1967, this document has been prepared in order to feedback to National Civic Education Committees the present state of the work and also to make further requests to National Committees.

This document is in six sections as follows:

1. A Working Definition of the Content and other Domains of Civic Education - (This is a result of an analysis of the National Committee reports submitted to the International Committee

2. The International Master Grid for the Content and Abilities of Civic Education: A Graphic Approach to Blueprints for Testing the Target Populations
   a) The Master Grid
   b) Topics by Target Populations:
      Population I
      Populations II and III
      Population IV

3. Sample Questions

4. Progress on Work Concerning Attitudes, Values, and Background Questionnaires
   a) Summary of Values
   b) The Measurement of Attitudes and Values in Civic Education

5. Hypotheses

6. Requests to National Centers
   (As will be seen from Section 6, one of the main requests is for items to be written. In this connection special attention should be paid to Sections 2 and 3).

We have documented the fact that civic instruction is given in varying degrees under various titles in all countries. It may be called history, geography, or social studies in England, Australia or Germany, for instance, and it may be given as a separate subject, such as in the United States, Iran, Sweden, Italy, or France for example. In Sweden, of course, "Civics as a school subject is dealt with in two ways: as a separate subject, and as a form of work in general, regardless of [the] subject being taught." (Marklund, Instruction in Civics in Swedish Schools, Page 5). In other words, civic education may permeate other subjects although, nominally, the subject being taught may be "economic," "social studies," or the like. Similarly, the committee fully recognizes the variety of civic instruction within countries, depending upon school type (private, public, comprehensive vocational, technical, academic, rural, urban, etc.). Allowances for this variance will be made in the content specifications, the testing instruments, the background information, and the interpretation of results.

The committee does not mean to imply that civic education is synonymous with the "social studies" broadly defined. Instead, within the social studies, there are aspects of geography, history, economics, and other studies which are closely akin to studies carried on in civics, citizenship, government, problems of democracy, international relations, current events, and other clearly identifiable civic education subject. The aspects of the social studies included in this study are those concerned with human political interrelationships on various governmental levels in organized societies.

By civic education, then, we mean that instruction which influences the political socialization of the child. That there are formal and informal influences on this education (family, school, peer group, mass media, etc.,) goes without saying. We shall attempt to sort out such sources and to weigh their influence. At the moment, however, the committee is concerned with identifying those phases of formal education which help develop knowledge, understanding, attitudes, and behaviors related to governmental affairs at all levels (local to worldwide). To repeat, those other institutions (economic, religious, family, etc.,) which influence the socio-political life of the student are certainly part of one's governmental education, are certainly important, and must be assessed. However, the committee has chosen to identify those school subjects at the various age levels having political science content and to make an over-all list of the major topics (e.g., democracy, constitutions, community, etc.,) covered in many or all nations.

1. A Working Definition of the Content and Other Domains of Civic Education

As planned in Bulletin 1 (December, 1966), the Civic Education Committee has examined the twelve nations' unique concepts of civic education. As we said there "After content analysis we intend to revise our present concept of civic education." The rationale statement and model concerning power, decision making and deference components, affective, cognitive, and behavioral domains and relationships from the level of selfhood to the world were present as a general *a priori* scheme of analysis for comparative educational research. This approach is well recognized and is familiar to students of comparative government.[1]

The committee has been most pleased with the many useful responses from the National Centers regarding their answers to our first four questions in Bulletin 1. These replies have aided us in redefining our concept of civic education. The two guiding questions for this redefinition were: 1) "Is education for citizenship in the curricula of all countries or not?" and 2) "Does a subject called "civics" exist in all of the countries engaged in this inquiry?"

We have conducted a content analysis of the documents received (including supplementary material on France and Israel). This *a posteriori*, or inductive approach, has had an effect on our hypothesized schema or picture of international civic education. Nevertheless, we had anticipated many of these results owing to the international composition of the committee and members' general knowledge of comparative educational systems.[2]

From the start, the civic education committee recognized the difficulty of defining the content, cognitive, affective and behavioral domains of the field. Thus, it was in November, 1966 that the committee first changed its name from the "civics" committee to the "civic education" committee, before beginning work. This was necessary because the meaning of the word "civics" differs among and even within, countries. The committee conceived of its task as being the study of political education or instruction in govenment or political science in the twelve participating nations - whatever title may be utilized in the country itself (citizenship, civics, political education, social studies, political or constitutional history, occupational orientation, etc.)

---

[1] Macridis, Roy C., *The Study of Comparative Government*. New York: Random House, 1955 (pp. 34-62).

[2] See Civics and European Education. Strasbourg: Council of Europe, 1963 (pp.150).

For purposes of this study, the definition of civic education for a country may partially be determined at each age level by the appearance (or indication) in a content-knowledge category or a national statement or goal which signifies that that objective is significant for, and characteristic of, instruction at that age level. A subject which may be titled work orientation, for instance, may be characteristic of elementary school instruction in Sweden and Italy, but not in other countries; thus, this topic might be examined by national centers in those countries and no others. On the other hand, a topic like the United Nations may be part of every nation's civic instruction and will therefore be part of the final school year examination in every country, as well as in those earlier years where the goals, curriculum guides, national or other laws, and tests detail such instruction.

The civic education committee has compiled an over-all content domain description (attached) resulting from our international content analysis. We realize that certain content topics may be mentioned in national goal statements, but not taught, and others may not be mentioned, and yet taught. The real crucible will be found after the testing itself, the interpretation of test results and the identification of those influential controlling or explanatory factors (formal or informal) as a result of attitude scaling and background questionnaires.

Thus, the committee decided to identify major content areas and cognitive abilities (e.g., ability to comprehend and interpret political data), then to ask which aspects of power, decision making and deference apply to certain of the major content topics (e.g., central and national government), and finally to determine if this is important and appropriate for measurement in different countries at the various age levels. We recognize the need to be highly selective in sampling from the expansive universe of material which could be tested in terms of power, decision making, deference, content, ability, age level, and level of abstraction (self to world). As a first step, therefore, only the most representative content and abilities dimensions have been described for each age level and level of abstraction. Categorization of questions into overlapping clusters dealing with power, etc., will have to await the actual production of test and attitude questions. In other words, power, decision making, and deference will both become central themes for the organization of the actual international civic education content and in terms of the questions themselves.

2. The International or Master Grid for the Content and Cognitive Domains of Civic Education: A Graphic Approach to Blueprints for Testing the Target Populations.

After the content analysis revealed that the word civics (or "citizenship") was a meaningful educational term in all countries (but with varying degrees of currency) it was decided to look for the least common denominator of political content among all the participating countries. The committee examined various national definitions of civics, political education, or social studies and found each of them either too connotatively overgeneralized (even for an international study) or too country-specific in their denotative meanings. The search for the philosopher's stone, the perfect and all encompassing definition suitable to all national centers, was soon aborted, however. The present definition was labelled both sufficient and necessary. Instead, the committee looked to the content (both overt and covert) of political instruction as defined by various countries, and selected those elements of the various social studies curricula which had political relevance (political geography, government economic policy, political history, etc.) Thus it was that one master content-abilities grid was prepared from the country-by-country analysis. Specific goals for age levels and for the attitude area as a whole (discussed below in Section 4) were also prepared.

2a. The Master Grid

The following grid is a general summary of the topics and abilities to be tested, as a result of our analysis of the national committee reports.

No cell weighting has been entered into the master grid but in general it is expected that at the lower age levels the amount of application and synthesis test items will be at a minimum, and perception and conception will be emphasized. By the upper age levels the number of application and synthesis items will comprise a larger proportion of the examinations. This plan fits in, it is assumed with both the psychological ability levels and the content level of abstraction for the age groups. This is to say that peer group, family, school, local community, etc., will be stressed at the lower age levels and national and international areas at the higher age levels. It is hoped that a detailed weighting for each cell can be arrived at over the summer of 1967 as a result of detailed analyses of national reports.

		Perception I	Conception II	Application III	Syntheses IV
A)	Nature of Citizenship	/	/	/	/
	1) Concepts of sovereignty, polity, stage, political communities, nation, loyalty, democracy, etc.				
	2) Rights, duties, obligations, and responsibilities (e.g. minority rights and majority rule)				
	3) Individualism, freedom, authority, rule of law, cooperation, leadership, participation, etc.				
B)	Political Processes and Institutions	/	/	/	/
	1) Local, departmental, state (or intermediary levels), national and comparative government (including political-historical development)				
	2) Political processes-public opinion, political parties, pressure groups, elections, etc.				
	3) The legislative branch - law making, constitutions, etc.				
	4) The executive branch, including the bureaucracy (civil service)				
	5) The judicial branch and courts.				
	6) Governments and social services, e.g. insurance, working conditions and unemployment, social security, welfare, education, medical care, health, roads, housing, personal security, etc.				
	7) Foreign policy, national defense; and international relations and organizations (e.g. balance of power, collective security, UNO, the emerging nations, etc.)				

	I Perception	II Conception	III Application	IV Syntheses
C) Economic Processes and Institutions	/	/	/	/
1) Production, distribution, and consumption.				
2) Industrial organizations, labor unions, and agriculture.				
3) Money and banking, the market, inflation, and public finance (taxation).				
4) Insurance, demography, natural resources, the world of work, and occupations.				
5) Government and the economy, economic trends, standard of living, etc.				
D) Social Processes and Institutions	/	/	/	/
1) The political socialization process (including family, school, community, peer groups, etc.)				
2) Communications, mass media, transportation, and advertising.				
3) Recreation and leisure time, group action, traffic, crime, welfare and service institutions.				

Topics

The topics as given in the master grid cover all age ranges and the list is inclusive. As will be seen from the three lists of topics for the different target populations given in the next section only some parts of the grid are used, particularly for the 10 year old population, but these parts are often expanded.

Note: A detailed international frequency analysis will be calculated in the summer of 1967 with specific illustrations of each content division drawn from the national reports which specifically identify (in statement form) the content, level of abstraction, etc., for each age level.

Abilities

Recognizing the great diversity in civic education content in the various nations, the committee decided to keep the recall, selection, and recognition level at a minimum. Certain facets of current events, international relations and foreign policy, and international organizations can safely be measured on the higher recall, selection, and recognition levels. However, the following categories of cognitive-abilities dimensions (which actually encompass or subsume the recall, selection, and recognition of factual information) were agreed to:

I. Perception: The ability to comprehend and interpret political data and information concerning power, decision making, and deference and to gather, receive and extract from such political data. The student: comprehends and understands political information; has an insight into the meaning of data that goes beyond mere recall; illustrates skills of translation of political symbols (e.g. political party symbols); can summarize and interpret a publication; demonstrates the ability to extrapolate trends.

II. Conception: The ability to analyse and evaluate political situations, problems, and information concerning power, decision making, and deference. The student: determines and applies the criteria to choose among meaningful alternatives, to make a political judgement about excellence in political writing to estimate progress toward social goals, to weigh different plans of action, or to determine the best criteria for distinguishing among democratic and autocratic political systems.

III. Application: The ability to apply the results of political analysis and evaluation to political problems concerning power, decision making, and deference. The student: demonstrates the skill to apply theories or abstractions discussed in a study to a specific or concrete political event (or events) and proves his ability to handle statistical information and other social science tools of analysis useful for the study of public problems.

IV. Synthesis: The ability to synthesize political information concerning power, decision making, and deference. The student: demonstrates his ability to distinguish fact from opinion; to compare political subjects, systems, and generalizations in order to produce an original hypotheses or plan of action; to generalize with insight; to derive a set of abstract relationships; and to test hypotheses or modify them on the basis of new informations.

The selection of this list of major cognitive areas was done with the full knowledge that much of the extant testing materials in various nations is rendered useless because of its factual recall or recognition orientation. Most of the materials which are appropriate to measure these areas of perception, conception, application, and synthesis, for instance, would require open ended or brief essay questions they may not be measured in this study. However,it is expected that the examples provided below and those which are now available in various national examinations (in Sweden, Australia, and the United States, for instance) will provide some of the material needed for the pretests to be constructed in late August, 1967.

2b. Topic by Target Populations*

In this section a figure is placed in parentheses after each major head. This indicates the rough weighting to be attached to each major head and therefore gives National Civic Education Committees an idea of the proportional number of questions to be produced for each sub-section (4= major emphasis; 3= considerable emphasis; 2= some emphasis; and 1= little emphasis).

(i) Population 1 (10 year olds)

A. Nature of Citizenship (4)

1. Duties and Responsibilities:
    - a) Respect for others and their property (including property)
    - b) Politeness (kindness)
    - c) Helping others (younger, old and handicapped persons)
    - d) Law observance, propriety, rules, conventions
    - e) Participation

2. Rights and freedoms (to receive a-d above)
    - a) Existence and survival
    - b) Worth of the individual
    - c) Member of society
    - d) Right to persuade, speak, reason, (a-d above), and to participate

B. Political Processes and Institutions (1)

1. National symbols and songs
2. Local government
3. Central government
4. International organizations (UNO, Red Cross)
5. Peace and conflict

C. Economic Affairs (2)

1. Education, work, and recreation
2. Consumership and saving
3. Taxes
4. Property
5. Advertising (mass media)

---

* For definition of Target Populations see Bulletin 2 (May 1967) pp.2-3

D. Political-Social Processes and Institutions (3)

1. Family life
2. School life
3. Peer groups
4. Communications (mass media) and transportation
5. Housing, traffic planning, armed forces, hospitals, public security, welfare, and service institutions

(ii) Populations 2 and 3

A. Nature of Citizenship (2)

1. Concepts of sovereignty, polity, state, political communities, nation, loyalty, democracy, etc.
2. Rights, duties, obligations, and responsibilities (minority rights and majority rule)
3. Individualism freedom, authority, rule of law, cooperation, leadership, participation, etc.

B. Political Processes and Institutions (4)

1. Local, departmental, state (or intermediary levels) national and comparative government (including political-historical development)
2. Political processes-public opinion, political parties, pressure groups, elections
3. The legislative branch-law making, constitutions, etc.
4. The executive branch, including the bureaucracy (civil service)
5. The judicial branch and courts
6. Governments and social services, e.g., insurance, working conditions and unemployment, social security, welfare, education, medical care, health, roads, housing, personal security, etc.
7. Foreign policy, national defense, and international relations and organizations (e.g., balance of power, collective security, UNO and the emerging nations.)

C. Economic Processes and Institutions (3)

1. Production, distribution, and consumption
2. Industrial organizations, labor unions and agriculture
3. Money and banking, the market, inflation and public finance (taxation)
4. Demography, natural resources, the world of work and occupational choice
5. Government and the economy, trends, standard of living, etc.

D. Social Processes and Institutions (2)

1. The political socialization process (including family, school, community, peer group, etc.)
2. Communication, mass media, and advertising
3. Recreation and leisure time, group action, traffic, crime, etc.

(iii) Population 4

A. Nature of Citizenship (3)

1. Basic political concepts of sovereignty, polity, political communities, nation, loyalty, etc.
2. Legal rights, freedoms, duties, obligations, and responsibilities, e.g., minority rights and majority rule
3. Individualism, freedom, authority, rule of law, cooperation, leadership, etc.

B. Political Processes and Institutions (3)

1. Local, departmental, state (or intermediary levels) national and comparative government (including political-historical development, separation, division, and interrelationships, of powers)
2. Political processes and policy formulation - public opinion, political parties, pressure groups, and elections
3. Central government
    a) The legislative branch - law making constitutions, etc.
    b) The executive branch including the bureaucracy (civil service)
    c) The judicial branch and courts
4. Governments and social services, e.g., insurance, working conditions and unemployment, social security, welfare, education, medical care, health, roads, housing, personal security, etc.
5. Foreign policy, national defense, and international relations and organizations (e.g., balance of power, peace and conflict, collective security, UNO, underdeveloped nations, etc.)

C. Economic Processes and Institutions (2)

1. Production, distribution, and consumption
2. Industrial organizations, labor unions, and agriculture
3. Money and banking, the market, inflation and public finance (taxation)
4. Insurance, demography, natural resources, the world of work, and occupations
5. Government and the economy, economic trends, standard of living, etc.

D. Social Processes and Institutions (1)

1. The political socialization process (including family, school, community, peer group, etc.)
2. Communications, transportation, mass media, and advertising
3. Recreation and leisure time, group action, crime, traffic, welfare, service organizations, etc.

* * *

3. Sample Questions

The following tentative, illustrative items have been suggested for Population 1 (10 year olds). It is hoped that they will be useful to members of national committees when they come to write items. On the left hand side of each item is a code indicating the content (A, B, C, or D in the topics by Target Population's chart) and the ability (I, II, III, or IV - see master grid). Under each item in parentheses is an indication of the general categories (power, decision making, and deference) identified for organizational sub-scores. It should be noted that these items are only a few of those reviewed by the international committee, but it is hoped that many better items will be written. (it is hoped to send out illustrative items for a) Populations 2 and 3 and b) Population 4 in the very near future.)

1. In your country the government makes all children attend school for so many years. Why does the government do this?
D2, (A) Because fathers and mothers who work cannot take care of their children.
(B) Because it is dangerous for children to play on the streets all day.
* (C) Because children must be taught to be educated members of your society.
(D) Because your teachers need jobs.
(E) Because your teachers and the school principal (or headmaster) have decided so.
(Power, Decision Making)

2. Stimulus material: Line drawing of a policeman apprehending a rich man and a vagrant who have picked flowers in a public park.

What is happening in this picture?

Ald, (A) Two people are hurt and another man is helping them.
I (B) Three people are playing a game in the public park.
(C) Three people are picking flowers in a public park.
* (D) A policeman is taking two men to jail for picking flowers in the public park.
(E) A policeman is hurting two people who have done nothing wrong.
(Power)

3. What does this picture tell your about the law?

A1d, (A) Only policemen can pick flowers in the public park.
(B) Only policemen obey the law.
(C) All taxpayers should be able to pick flowers in the public parks.
* (D) The law is for all people, rich and poor alike.
(E) Everyone should learn to obey the law because a policeman may be nearby.
(Power)

4. Stimulus material: Line drawing of a courtroom trial with bench, robed judge or judges, etc.

This is a picture of a (an)

B2a (A) school
3, (B) office
I (C) library
* (D) trial (or courtroom)
(E) hospital
(Power)

5. Which of the following things does your government make, and that no one else in your country is allowed to make?

B3,
IV

I. Postage stamps and money
II. Passports and drivers licenses
III. Bank check and bonds
IV. Flags and guns

* (A) I and II only
(B) II and III only
(C) III and IV only
(D) I, II, III, and IV
(Power)

6. Stimulus material: Line drawings of each of the following occupational groups. Which of the following people work for the government? Circle (a) if the person works for the government, (b) if he does not, and (c) if you do not know whether or not he works for the government.

C1,				
I	a	*b	c	A milkman
	*a	b	c	A public school teacher
	*a	b	c	A soldier
	*a	b	c	A congressman, M.P., etc.
	*a	b	c	A policeman
	*a	b	c	A judge
	*a	b	c	A postman

(Power

7. Stimulus material: Line drawing of Italia Con Corona Turrita, Brittania, Uncle Sam, Adler, etc., or picture of national paper currency with an appropriate national symbol.

B1,
I ________________________is dressed like he (she) is because this
* (A) figure stands for a symbol of your nation

(B) is the way people dressed long ago
(C) shows that your country belongs to the United Nations
(D) shows that the money was printed by the government
(E) is patriotic
(Deference)

8. We must put stamps on our mail in order to
C2, (A) avoid paying a fine
II (B) help stamp collectors in other countries
* (C) pay the government for delivering mail
(D) show that we are good citizens
(E) prove we know the mail regulations
(Power)

9. Peter usually walked to school, but today he was riding with his father because he had stayed in bed too long. He told his father, "Hurry, father, you are driving too slowly. I will be late for school and that is against the rules." "Sorry, son," said his father. "Thirty-five kilometers (miles) per hour is the speed limit, whether you are late or not."

A1d, What is the best reason why Peter's father should obey the speed
III limit?
(A) The police may see him
(B) The law sets the speed limit
* (C) The danger of an accident is less
(D) The people who speed are punished
(E) The car engine will be harmed if it goes too fast.
(Decision Making)

10. Who made the law about the speed limit?
A1d, (A) The drivers
I (B) The police
(C) The judges
* (D) The government
(E) The automobile companies
(Power)

11. What must every citizen do even before he becomes a voter?
* (A) Pay taxes if the government requires it
C3, (B) Go to religious services regularly
I (C) Join a political club or party
(D) Go on to a college or university for further education
(E) Memorize the Constitution and the national anthem
(Power)

12. Which of the following can best be determined after a school class votes on it?
A2d, (A) Deciding the best way to make a sick classmate well again
III (B) Deciding whether a small chicken which someone has brought into class is a male or female

(C) Deciding what the price of bread in the school lunchroom should be
(D) Deciding which student in the class has the best examination paper
(E) Deciding who shall be the class representative or officer
(Decision Making)

13. Which of these students shows most clearly that he has not been a good citizen in the classroom?
(A) One who has not done his homework
* (B) One who has stolen something from another classmate
D2, (C) One who has stumbled and broken a pane of glass
Ala, (D) One who has been responsible for his team losing in a class game
(E) One who has misspelled some easy words
(Deference)

14. The children in a class had made their own rules. They agreed to do the following things:
(A) To put up their hands when they wanted to speak
Ald, (B) To keep themselves and their desks tidy
IV (C) To keep to one side on the stairs and in the corridors
(D) To stop others from throwing stones, sticks, or anything that might hurt someone
(E) To speak politely and in a courteous and friendly manner to others
(F) To do their homework regularly
(G) To eat in a mannerly way and to keep the lunch room clean

What are the letters of those rules which will protect pupils from accidents? C & D

What are the letters of those rules which will make you a good citizen of the school if you follow them?
All

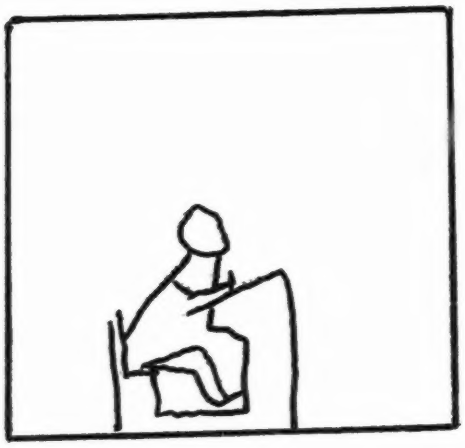
(Decision Making)

15. Rules are made by different people for different reasons. Some rules are made by governments for all people. Which rules below are made by the government? Circle (a) if they are made by the government, or (b) if they are not, or (c) if you do not know who makes the rules.

D2,
Ald,
IV

*a	b	c	Traffic regulations
a	*b	c	Regulations for playing sports
a	*b	c	Lunchroom regulations
a	*b	c	Rules for table manners
*a	b	c	Voting regulations

(Power-Decision Making)



16. Why does Johnny have to study in school?

D1, (A) Because otherwise his parents will not take him on vacation

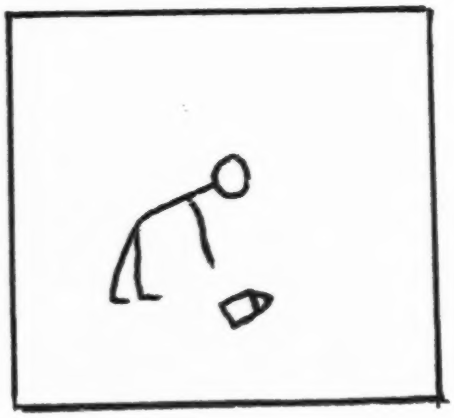
I * (B) Because it is his job to learn many things which will be useful to him

(C) Because he is the best student in the class

(D) Because the government makes him go to school

(E) Because the other children study

(Decision making)



17. Suppose that Johnny finds a wallet on the ground, what should he do?

(A) Give it to a poor man
(B) Keep it for himself
(C) Give it to a policeman
(D) Divide the money among his friends
(E) Take part of it for himself and give the rest back to the owner

(Deference)

4. Progress on Work Concerning Attitudes, Values and Background Questionnaires.

During the Rome meeting a separate subcommittee was established to work on the areas of attitudes and behaviors in civic education. This subcommittee will be working in conjunction with the international committee. Attitude scales, background questionnaires, and similar instruments will be produced conjointly. The subcommittee on attitudes (which will receive additional help from the IEA technical committee and from other experts in the field) was composed of Dr. A. N. Oppenheim of the London School of Economics and Mr. T. N. Postlethwaite, the IEA coordinator.

4a. Summary of Values

The following summary is the result of a survey of the national papers in order to detail the general values which appear time and again. (A more complete list with list of values is being compiled).

Democratic Values Stressed In International Civic Education Documents.

1. Tolerance of diversity and egalitarianism. The idea that all citizens are of equal value as human beings and that their differences are a source of interest and strength to the country.

2. Acceptance of interdependence and involvement: national, international (UN), and historical. The idea that no man is an island, and that people are connected with one another in groups, as members of states both at the international level and with their past.

3. Respect for the balance between the liberty of the individual, freedom of the press, assembly, worship, expression, etc., versus the rights of others. The ideals of equality before the law, civic rights, nondiscrimination, and self-restraint.

4. Attitudes about the use of and limitations on state power, accountability of elected representatives, voting, and political behavior.

5. Attitudes toward certain functions of the state in (1) providing services, social justice, welfare, security, planning, etc., (2) regulation of competing interests, e.g., antimonopoly legislation, trade unions, and consumer protection.

6. Acceptance of the ideal of progress and social change through informed participation versus the feeling of powerlessness; social concern, and a rational and critical approach to social problems; work for the common good; incorruptibility; informed public opinion.

7. Belief in the pacific regulation of conflict; peace; nonviolence, group decision making processes; majority versus minority rule.

8. Belief in the need for participation in school, family and group decisions.

Other Values, or Those Less Emphasized or Only Emphasized in One or a Few Countries.

9. Preparation for life, familiarity with the functioning of modern society; job orientation, leisure, and vocational opportunities, and the acquisition of new knowledge independently.

10. Love and loyalty to country; solidarity, reasoned patriotism.

11. Moral principles, e.g., truthfulness, integrity, responsibility efficiency on the job, preserving one's health; thrift, saving, and enterprise.

12. Cultural and aesthetic values.

13. Fear of God and Christian charity, religious values.

14. Unification of Europe, internationalism.

15. Problems of leisure.

16. Deference to authority.

Of the values in the second section only numbers 9 and 10 were frequently mentioned.

After having summarized the values mentioned in national committee reports, Dr. Oppenheim prepared a document concerning the measurement of attitudes and values in Civic Education. This document is reproduced below. It suggests a special pilot study to construct some of the instruments. It is hoped to begin this pilot study in the near future.

4b. The Measurement of Attitudes and Values in Civic Education.

by

A. N. Oppenheim

Attitudes as Dependent and as Independent Variables

1. For the sake of completeness it should be pointed out at the beginning that there will be a number of attitudinal measures included among the predictors of all subject achievement. In the area of Civic Education such predictors might include, for instance, some indication of parental attitudes, or the type of discipline favored by the teacher. The attitudes are to be measured in this way, i.e., as independent variables. A list of these will provide the subject of a separate document. What follows is primarily concerned with attitudes as dependent variables; in other words, these are the attitudes and values which we hope (or expect) to find developing in children, partly as a result of school teaching processes and partly as a result of various aspects of socialization generally.

2. We recognize that there are close links in this domain between perception, knowledge, values, and behavioral tendencies, with influences going backwards and forwards - such that an absorption of knowledge is partly selective in line with pre-existing attitudes, but also that further knowledge may influence attitudes and values. If we seem to separate attitudes from knowledge, this is purely done for purposes of convenience and should not imply a misleading dichotomy.

The Contents of the Country Documents.

3. At the outset there was no indication as to whether or not there might be a common core of values and attitudes which are present or implied in the teaching programs of all or most countries involved in the research. It was therefore necessary to conduct a fairly detailed content

analysis of the country documents in order to ascertain the various values that were being emphasized, at different age levels. The results of this analysis (see previous document) tended to show that all countries concerned emphasized a broad area of democratic values. In 4a these values have been specified under 8 headings, which are fairly arbitrary in their divisions. We have tried to indicate the degree to which each of these in turn is emphasized in each country, according to the documents supplied to us. Of course, we recognize that different countries have interpreted their task differently, and that the documents do not necessarily always reflect accurately what is classroom practice. Moreover, in the field of attitude transmission much is done by implication rather than by direct tuition. Much depends on the way in which factual material is taught, as well as on the actual content.

4. In addition to this broad area of 8 major points covering democratic values, we found that a number of other attitudes were mentioned (by one or more countries) as having some emphasis in the curriculum. (See list of other values 9-16.) For instance, fear of God and Christian charity were mentioned in one country, problems of leisure were mentioned in another, and so on. Possibly some of these additional values could be encompassed in the broad framework of democratic values, while others might have to remain separate or might otherwise become the subjects of country-specific attitude measures.

Approaches to Measurement

5. It now becomes incumbent upon us to stand back from the material and to ask ourselves: "How can these detailed statements derived from the curricula be linked with existing knowledge about democratic attitudes?" It would, of course, be possible to go ahead and produce 10 or more attitude measures, each covering a sub-area of democratic values. But it might be possible, and perhaps more valuable, to develop one or two broad, underlying measures. The advantages of so doing would be threefold:

i) it would reduce testing time considerably

ii) it would link the measures in this project to existing psychological knowledge, and make a contribution to it.

iii) the overt contents of such underlying measures would be somewhat less "political" in content, thus making it easier to secure permission to have them administered to school populations

6. Over and above these three advantages there is another, which is specifically of relevance to the aims of IEA; namely that it may be possible to develop these broad, underlying measures in a form that will make intercountry comparisons possible and valid. If this were so, then, leaving aside for the moment problems of translation, etc., we could hope to end up with country measures or country profiles that would be comparable. The results of the inquiry might hope, then, to show two things:

i) the extent to which pupils at different age levels in each country had progressed along measures of democratization

ii) the extent to which each country, in turn, had achieved, exceeded, or fallen short of its own stated aims in this area.

7. It is worth pointing out that we expect to determine if the school is or is not one of the major influential factors having impact on the development of these values in the child. We may find that open socializing processes will, in this domain, come more to the fore than elsewhere. In this study we expect to find that the influence of the school will vary, by degrees, from country to country and from age group to age group even within a country. Ultimately it would be helpful to governments and other readers of our reports to know to what extent democratic values were developing in the children, through whatever means, and which are irrespective of the school curricula.

8. The area of democratic-authoritarian values is one that has long been the subject of attention is social psychology and other social sciences. Many early studies were brought into sharp focus in the 1940's by the work of Adorno and others on the authoritarian personality. The F-scale, which was a result of this massive work, covers the type of underlying variables that were referred to above. It includes, for instance, the tendency to use stereotypes and black and white thinking, tendencies to projectivity, emphasis on power and strong leadership, fatalism and so forth. The broad stream of research which followed this study has shown that scales of this kind are closely related to the more detailed democratic values in civics (which are mentioned in 4a) for adults. There have been studies of

authoritarianism and rigidity in children - mostly attempting to link the development of such attitudes to various types of regimes in the home but no generally accepted and validated F-scale for children exists at the moment.

9. In this context mention should be made of later work such as that of Rokeach on the open and closed mind, and the work of Eysenck on basic social attitudes. Neither of these research workers has produced scales directly applicable to children, however, Eysenck's work has two possible advantages to our study:

i) It presents a broad constellation of political and social attitudes, including authoritarianism, within a two-dimensional framework of: (a) radicalism/conservatism, and (b) tough mindedness/tender mindedness. Within this framework high scorers on the F-scale find themselves located in the tough-minded, conservative quadrant.

ii) The work has been repeated a number of times in various European countries, and the same factor structure has emerged; this means that while details of items and translations may vary, the underlying dimensions are roughly similar, and scoring can take place on a basis which is comparable across countries.

Proposed Pilot Study

10. It is not possible to ascertain the nature of such underlying factors in children without conducting a fairly substantial pilot study. Mounting such a study would probably be a separate exercise to be conducted by the technical committee. It should not be assumed that attitude scales are necessarily the best way of approaching this problem; other types of items should be tried, including projective techniques (if they can be made scorable). The detailed planning of the pilot study would have to be considered, but a rough plan might suggest the following steps:

i) the construction of an item pool or set of measures directly derived from the points in 4a above.

ii) additions to the pool derived from the literature on the use of F-scales with children, measures of rigidity, stereotyped thinking, and basic social attitudes such as those used by Rokeach and Eysenck, and possibly others

iii) translation and administration of this pool to groups of children at different age levels in 3 or 4 selected and contrasting countries

iv) a factor analysis or component analysis which would enable us, hopefully, to identify a very small number of basic underlying variables, common to the different countries and age groups, and scorable by factor loadings

v) preparation of the final scales

11. It is also worth mentioning that the conduct of this experiment is a major exercise in itself. It would, however, have all the advantages stated above and, in addition, make an important contribution to socio-psychological literature and knowledge.

Other Values

12. Decisions will have to be made about measurement of the values mentioned under 9-16 in 4a. Some of these may already be subsumed under other democratic values; others may have to be measured separately, and still others may fall by the wayside.

13. We also have to bear in mind some broad mental attitudes which have been suggested to us by others, such as:

i) tendencies to stereotyped thinking; (this is easily subsumed under authoritarianism and rigidity)

ii) measures of self-esteem, as a student and as a human being (like other attitude measures, this in particular could be both a dependent and an independent variable)

iii) measures of need-achievement (This refers to the research of David McClelland who has related broad measures of achievement-motivation in children and in adults to the degree of achievement-oriented content of mass media to a country's social and economic achievements, etc.)

5. Hypotheses

National Committees will have seen from Bulletin 2 the formulation of composite hypotheses being undertaken by IEA. The variables to be entered into composite hypotheses can be gleaned from the consideration of a series of shorter hypotheses. In November, 1966, Sixteen Marklund of Stockholm, Sweden wrote a short memo "Concerning Hypotheses" which will be of interest to National Committees. This is reproduced below.

Concerning Hypotheses

by

Sixten Marklund

A. The outcomes of civic education can be described and evaluated in many different ways and according to various dimensions: It is of prime importance at the beginning of such a study to obtain a rough indication of what civic education is or at least of what it is thought to be. The IEA civic education committee therefore needs information from participating countries about how they define civic education. Three factors are of main importance:

1. Objectives of civic education
2. Content of civic education
3. Subject organization in relation to civic education

B. As far as pupil achievement is concerned, there seems to be three main areas of interest: knowledge, attitudes and interest in the subject. Continuing the numbering from Section A they can be broadly sub-divided as follows:

4. Knowledge of facts, total mean
5. Knowledge of facts, profile of subscales
6. Knowledge of facts, dispersion in scales and subscales
7. Attitudes (feelings, etc.) in human relations: family
8. " " : peers
9. " " : school and teachers
10. " " : certain groups (minorities, etc.)
11. " " : police and other public officials
12. " " : international affairs
13. Pupil interest in civic education.

For all of these domains evaluation instruments must be constructed with which quantitative data can be collected and subsequent analyses carried out. The test and scale items must concern important concepts and features of human and political relations which are common to all countries (e.g. power, deference, freedom and decision making).

C. The domains of variables thus arrived at (4 - 13 above) can be treated as dependent variables. The independent variables should, of course, be detailed descriptions (in measurable terms) of the points 1, 2, and 3 mentioned in Section A as well as variables such as the following:

14. Amount of civic education teaching at different age levels

15. Percentage of pupils of an age group taught civic education at different levels

16. Categories of teachers, categories and information on their pre-service and in-service training

17. Methods of teaching in civic education

18. Tests and examinations in civic education

19. Types of civic education in different school types, e.g., comprehensive, selective, academic, etc.

20. Differences in civic education for boys and girls.

D. This list is far from complete and the variables can, of course, be made operational in a much more detailed and precise form. Further pupil and teacher background variables can be added such as those used in the IEA mathematics study.

E. More information in each of these areas must be collected. This data collection will constitute a necessary first phase in the civic education study. The generating and developing of hypotheses is strongly dependent on what this first phase yields.

The hypotheses will in general refer to a great number of inte-correlations between these variables. Differences in measurements will be found within countries as well as between.

F. The following are, therefore, only a few examples of hypotheses which may be tested in the study. They illustrate suggested types of hypotheses rather than final hypotheses:

1. Pupil knowledge of governmental affairs is higher when civic education is given as a separate subject than when it is an overall aspect of a number of subjects.

2. The pupils' interest in participating in social activities is positively related to their interest in governmental affairs.

3. Pupils' participation in extra-school activities (clubs, games, etc.) is related to their perception of authority relationships within the family.

4. The more a curriculum in civic education stresses a functional approach (as opposed to a structural approach) the greater will be the variability of the knowledge-profile pattern.

5. The variability of knowledge of international relations is positively related to the degree a curriculum in civic education stresses a structural approach.

6. The pupils' interest in the subject of civic education is higher when instruction is given in short, concentrated courses than when instruction is spread out over a whole year (or more).

7. Pupil difference between age 13 (±1) and age 17 (±1) in attitudes towards participation in school activities is positively related to the type of school organization which promotes such activities.

8. There is a negative correlation between the pupils' knowledge of international organizations and their thinking in stereotyped terms.

9. Etc.

6. Request to National Centers

1. We would like to remind you of the information requested for Phase B (see Bulletin 1, pages 83-84, items 5 - 20). (Some National Centers have already answered some of the questions posed for Phase B in their Phase A reports. Other centers have stated that they have not the resources to conduct the analysis needed to answer certain questions. If such information is not readily available, the expert or professional judgments of the national civic education committees will suffice for answering the specific query.)

2. (a) We would like to have further information on what political content is taught at specific age levels to specific portions of the school populations.

(b) At the same time we would appreciate receiving selected outlines of the political material (from textbook tables of contents, curriculum guides, governmental syllabi, etc.) taught to each age group. We do not want any more than a summary of geography, sociology, economics, history, etc., material which is taught unless this material has a political content, e.g., political geography, public opinion polling, international political history, public finance, etc.

(If you require any elucidation of what is requested in 2 a and b please write to Dr. Russell Farnen, Educational Testing Service, Test Development Division, Princeton, N.J. 08540, U.S.A.

3. (a) Many items are required for try-out, to this end we would like as many examples of currently used examinations with certain items categorized as to content, ability, age level, etc. in terms of the grid supplied in Section 2 above.

The international committee has (and intends to continue to do so) reworked examinations and questions submitted by national centers. However, the types of questions sought for this international survey are sufficiently unique as to be scarce in the literature. National centers are encouraged to use appropriate questions which appear in the other national reports they have received as prototypes for such questions. Centers are also welcome to adapt those sample questions which they find are appropriate to their nations or to modify them into a format suitable for use in their own nations. As these questions are written, the international character of these tests much also be kept in mind. Those national questions which may be used in several countries are the most valuable. The use of stimulus materials (maps, charts, graphs, cartoons, line drawings, quotations, etc.) is to be encouraged in this way the examinees will have some measure of common verbal or nonverbal material on which to base their responses.

4. We would be interested in receiving the national centers comments on the materials presented to them thus far (grids, sample items, rationale, etc.) The same invitation applies to the material on attitudes presented below.

5. All materials produced should be sent to the IEA coordinator in Hamburg before July 30, 1967. The next meeting of the International Civic Education Committee will be held at the end of August, 1967.

We thank you for your help.

Russell Farnen

APPENDIX H-1

Civic Education Content Specifications -
Cognitive Domain 8/29/67 (Hamburg, W. Germany)

	Simple	Complex	Abstract	Age Levels			
	I	II	III	10	14-15	18	18 elite
A. Fundamentals and Nature of Citizenship - Citizenship concepts (definitions).	/	/	/	/	/	/	/
1. Sovereignty, polity, state, government, political communities, nation, loyalty, patriotism, nationalism, legitimacy, authoritarianism, etc.							
2. Rights, duties, obligations, responsibilities, and power (e.g. minority rights and majority rule), etc.							
3. Individualism, freedom, authority, rule of law, cooperation, leadership, participation, democracy, amity, trust, equality of opportunity, etc.							
B. Political Processes and Institutions	/	/	/	/	/	/	/
1. Constitutions							
2. Historical (Social-Political-Economic) development, including political theory, separation of powers, etc.							
3. Local, departmental, state, province (or intermediary levels) as compared with central government, etc.							
4. The national legislative branch, law making, etc.							
5. The national executive branch (including the bureaucracy, civil service), the administrative branch, and the Cabinet, etc.							
6. The national judicial branch and courts.							
7. Political processes and institutions, public opinion, political parties, pressure groups, elections, political coalitions and decision making, etc.							
8. Foreign policy, national defense and foreign affairs, etc.							
9. International relations and organizations, cross-cultural, nonwestern, comparative politics and governments (e.g. unitary and federal systems, parliamentary and presidential systems, etc.)							

APPENDIX H-1 (cont'd)

				Age Levels			
	I Simple	II Complex	III Abstract	10	14-15	18	18 elite
10. Governments and social services e.g. insurance, working conditions and unemployment, social security, welfare, education, medical care, health, roads housing, personal security, etc.							
C. Economic Processes and Institutions 1. Government and the economy, economic trends, standard of living, foreign aid and trade, technology, etc.							
2. Industrial organizations, labor unions, and agriculture.							
3. Money and banking, the market, insurance, inflation, and public finance, (taxation).							
4. Demography, natural resources, the world of work, and occupations.							
	/	/	/	/	/	/	/
D. Social Processes and Institutions 1. Group interactions between the individual, peer group, family, school and community, interdependence, social interchange, and "socio-political" man.							
2. Communications, mass media, transportation, advertising, etc.							
3. Recreation and leisure time, group action, traffic, crime, welfare, and service institutions.							
E. Methods of the social studies disciplines, and Interdisciplinary Approaches i.e. How do the various social sciences study all the above, e.g. the economist, the historian. the political scientist, etc.	/	/	/	/	/	/	/
1. Methods of the Social Sciences							
F. Critical thinking and problem solving							
G. Current events							

Note: Factors such as current events, the problem solving method, and the like are meant to cut across all of the above areas (A-E) but are categorized as F and G for test assembly purposes.

6

Summary of Values Mentioned in National Reports (Rome, April, 1967)

DEMOCRATIC VALUES	Germany	Finland	Iran	Italy	Sweden	UK	USA
. Tolerance of diversity; legalitarian attitudes; the idea that all citizens are of equal value as human beings and that their differences are a source of interest and strength to the country.	xx	x	xxx	x	xx	x	xx
. Interdependence and involvement; national, international (UN) and historical. The idea that no man is an island; and that people are concerned with one another in groups, as members of states, at the international level and with their past.	x (Hessen)		xxx	x	xxx	x	xx
. Liberty of the individual, freedom of the press, assembly, worship, expression, etc., versus rights of others. Equality before Law. Civil rights. Self-restraint.	x xx	x	x	x	x	xx	xxx
. The use and limitations of State power, accountability of elected representatives; voting and political behaviour.	x	x	x	x	x	xx	xx
. Functions of the State in (i) providing services, welfare, security, etc., planning (ii) regulation of competing interests, etc., anti-monopoly legislation, trade unions, consumer protection.	x	x	x	x	x x	xx x	xx
. Equality before the Law; non-discrimination, equality of opportunity; social justice; non-discrimination	x		x		x	xx	x
. The idea of Progress through informed participation vs. feelings of powerlessness; social concern, and a rational and critical approach to social problems; work for common good, incorruptibility; informed public opinion.	x x		xxxx	x	x xx	x xx	xxxx
. Education for citizenship - is it in the curriculum or not?	x	x	x	x	x	x	xx
. Pacific regulation of conflict; group decision-making processes; majority vs. minority; peace; non-violence.	(Too little) x					?	x

DEMOCRATIC VALUES (Cont.)	Germany	Finland	Iran	Italy	Sweden	UK	USA
10. Participation in school decisions	x		x			x	x
11. Preparation for life: familiarity with the functioning of modern society; job orientation vocational opportunities; the ability to acquire new knowledge independently; anticipating and coping with social change					xx	xx	xx
OTHER VALUES							
12. Moral principles, e.g. truthfulness, integrity, responsibility, efficiency on the job and preserving of own health; thrift, saving and enterprise	x	x	x	x			
13. Cultural and aesthetic values	x		x	x			
14. Love and loyalty to country; solidarity	x			x			x
15. Fear of God and christian charity (religious values)	x						
16. Unification of Europe	x			x			
17. Problems of leisure					x		x
18. Deference to authority							xx

Illustrative Items for Civic Education

Population I:

1. According to the rules of democratic politics, victory in an election consists of securing the:

    a. support of businessmen
    b. most qualified candidates
    +c. largest number of votes
    d. greatest financial contributions
    e. support of the minority party

This question applying to national political processes and institutions (II) asks the ten year old student to identify the democratic principle of majority rule. It is expected to be abstract for the ten year olds because of the selectivity required rather than the language level.

2. Stimulus material: Pictures of five flags the UNO flag, the international Red Cross or Red Crescent flag, the Olympic games flag, and the Boy Scouts and Girl Scouts flags labeled flags 1, 2, 3, 4, and 5. Which of these flags is the United Nations flag?

    +a. Flag 1
    b. Flag 2
    c. Flag 3
    d. Flag 4
    e. Flag 5

The recognition of the UNO flag by the 10 year old would show the beginning level of familiarity with international organizations (III). It is expected to be a relatively simple task for all countries involved including West Germany which is not a UNO member.

3. Why must we put stamps on our mail?

    a. to avoid paying a fine
    b. to help stamp collectors in other nations
    +c. to pay the government for delivering mail
    d. to show the government that we are good citizens
    e. to show everyone that we know how to mail letters

This question will be somewhat complex for 10 year olds. It is designed to test the child's understanding of why we must pay for governmental services (IV). There should be few difficulties with translation of this item since it is a fairly universal practice across the seven nations involved in the civic education assessment.

4. What is the most important duty of the police in a community?

a. to punish criminals
b. to hand over criminals to police centers
c. to put thieves in prison
+d. to help people and protect their lives and properties
e. to arrest people and take them to court

This question for 10 year olds tests social processes and institutions (V) and is complex for the 10 year old. The question asks for the most important duty of the police and each of the other distractors is either plausible or an actual practice of the police. The translation of the key words police and community should pose no problem in any country.

Population II

1. Which of these would most probably be the best citizen in a democracy? Someone who:

a. has good health and much money
b. wins many athletic contests
+c. reads about current events and votes in elections
d. is born into a good family and has many friends in important places
e. has had much schooling

This simple question for 14 year olds tests citizenship fundamentals (I) in terms of informal political participation. It is expected to be a simple question across countries since it deals with a statement of a democratic ideal which is commonly taught in the participating nations.

2. Which of the following correctly describes your country's Constitution?

+a. Federal and unwritten only
+b. Unitary and unwritten only
+c. Unitary and written only
+d. Federal, written, remaining powers lie with federal government
+e. Federal, written, remaining powers lie with state/provincial government

This is an example of an abstract question for 14 year olds dealing with the national constitution and the distribution of governmental powers. The key will vary by nation. e.g., Iran, Italy and the UK, will be b/ Sweden and Finland, c/ West Germany, d/ and the United States, e. It is categorized as national political processes and institutions (II).

3. Stimulus material: cartoons showing charter members and emerging nations on a seesaw with the U.N. building as the fulcrum. The emerging nations are weighing down their end of the seesaw.

   The cartoonist is saying that:

   - a. underdeveloped countries need United Nations help
   - +b. there are more members who have joined the United Nations than there are charter members
   - c. the emerging nations in the United Nations represent a larger total population than charter members
   - d. United Nations charter members must respect the military power of the emerging nations
   - e. underdeveloped countries do not take the United Nations seriously and they are only playing a game with the charter members.

This will be an abstract item for 14 year old students in all countries. It tests international topics (III) and requires the interpretation of graphic material as well as fundamental knowledge of the U.N.

Population IV (General)

1. Citizens of a democracy are free to choose only four of the following. Which one are thez NOT free to choose?

   - a. the religion they will follow
   - b. the job they will pursue
   - +c. the laws they will obey
   - d. the newspapers they will read
   - e. the groups they will join

This complex question for the preuniversity year (general) school population asks students to select the fundamental citizenship obligation (I) of law observance from a list of alternatives where there is freedom of choice.

2. Which of the following statements best support the argument that pressure (private political action) groups are legitimate instruments in the democratic political process?

   - a. they generally advocate what the majority of the people desire.
   - +b. they exercise the rights of petition and assembly
   - c. their internal organization and mode of operation are democratic
   - d. their concern is to promote the national interest
   - e. they are in closer contact with the general public than are government officials.

This complex question, categorized as national political processes and institutions (II), associates the activities of private interest or pressure groups with the democratic political process. The student is asked to apply the democratic rights of petition and assembly to such groups and to accord them legitimacy.

3. Only four of the following statements accurately describe democratic politics. Which one does NOT?

   a. politics is a process in which most of us take part at one time or another
   b. politics deal with vital issues of our time
   c. politics deal with the economic health of nations
   d. politics deal with settling conflicts among people, groups, and nations over the realization of goals
   +e. politics mainly deal with petty and sordid things such as how can propaganda be used to further general welfare programs

This abstract question regarding national political processes and institutions (II) asks the student to differentiate among four typical political activities and one which conflicts with usual political practices in democratic political systems, i.e., petty, sordid, and propagandizing activities.

4. The goals of the European Common Market (E.E.C.) are mainly being achieved by:

   a. exchanging armed forces in member nations when needed
   b. granting more self government to the member nations
   c. stopping any products from nonmember nations from entering member nations
   +d. eliminating tariff barriers among the member nations
   e. granting loans to underdeveloped member nations

This complex question on international political processes and institutions (III) asks the student to identify a primary means by which the common market is achieving its purposes, i.e., a free trade area among the "Inner Six."

Population IVS (Specialists)

1. Which of the following principles are essential today for the operation of a democratic political system?

   i. rule by the wise, rich, or just
   ii. political party opposition and competition
   iii. one man or one party rule

IV. Universal suffrage (regardless of sex, color, religion, etc.).
V. Periodic elections

a. I and II only
b. III and IV only
+c. II, IV, and V only
d. III, IV, and V only
e. I, II, III, IV, and V

In this simple question the student is asked to reject elitism, absolutism, or totalitarianism as democratic characteristics and to recognize the need for political competition, regular elections, and broadened suffrage requirements as elements of democratic theory and practice. The question is categorized as national political processes and institutions (II).

2. Which of the following is the best electoral system for democratic countries?

a. single-member district of representation
b. proportional representation of all parties according to voting strength.
c. double-ballot, single-member district of representation
d. suffrage restricted to secondary school and university graduates
+e. none of the above, because particular political systems vary from state to state

This complex question asks the student to identify democratic relativism for various political systems as well as to reject the restricted suffrage. He is expected to see that no one of these systems is inherently superior to any other and that, in fact, it is the way in which the system operates which is the most important criterion of political success.

3. Which of the following were established primarily to lower customs duties among member countries?

I. The World Bank
II. General Agreement on Tariffs and Trade (GATT)
III. Food and Agriculture Organization (FAO)
IV. International Monetary Fund (IMF)
V. European Economic Community (EEC)

a. I and II
b. II and III
+c. II and V
d. III and IV
e. IV and V

This simple question dealing with international organizations (III) asks the student to sort out the work of certain U.N. agencies from that of Common Market operations on the basis of a primary purpose, that of reduction in customs and tariffs. The same is true for GATT.

4. A "mixed" economic system is one in which there is:

   a. equal emphasis on manufacturing and agriculture
   b. both a national and a local tax leveled on the general public
   c. balanced government spending for civilian and military purposes
   +d. important economic decision making by both public and private agencies and individuals
   e. a movement from a rural to an urban society

This complex question asks students to select the classic definition of a "mixed" economy from among relevant, but erroneous distractors. The mixed economy, of course, is to be contrasted with the unregulated laissez faire economy on the one hand and the rigidly regulated economies of e.g. Communist China or the U.S.S.R. on the other. It is categorized as economic processes and institutions (IV).

5. Suppose that we know that since 1945 juvenile delinquency was increased greatly particularly in urban areas in a highly industrialized society which we shall call Country X. Which of the following conclusions can be drawn from this?

   a. peace has caused an increase in the crime rate
   b. young people in cities are basically more evil than young people in the country
   c. police are not as well trained today as they were formerly
   d. all young people are inherently more evil now than they were formerly
   +e. none of the above conclusions can be drawn

The abstaact question for the preuniversity specialist group is designed to test the student's knowledge of certain social processes and institutions (V), i.e., crime, its causes, the police, and the nature of contemporary youth. Since none of these explanations will endure critical analysis, the student is expected to reject all but the last alternative.

REPORT ON THE I.E.A. PILOT INTERVIEWS FOR CIVIC EDUCATION (AFFECTIVE DOMAIN) by

Margarete Harvey

This report is based on a series of free interviews with school children in England and Germany carried out in the early summer of 1967. Their purpose was the identification of the main dimensions to be included in the affective domain of the project and, if possible, recommendations for the methods to be used.

The procedure was as follows: during the interviews, which lasted between 60-90 minutes and were carried out on the school's premises, the child would be confronted with a variety of stimuli and asked to comment freely. The stimuli consisted of attitude statements used in earlier research, pictorial material, and introductory and/or follow-up questions by the interviewer. The stimulus material had been chosen to elicit responses relevant to the "Democratic Values" as listed in Document A., which was attached to CIV/RM.3. Deliberate effort was made to include varied approaches in order to attack the 'democratic values' from as many different angles as possible

Since there was a wealth of relevant material, roughly 1/8th was utilized in the course of each interview, so that all the stimulus material was tried out at least once in each of the six schools visited. Moreover, some of the stimuli were only used with one age group, others were excluded as soon as they proved to be totally inadequate for the purpose outlined.

The interviews were tape-recorded and then transcribed in some detail, listing the child's main concepts, the topics covered and suggestions for items. The following generalizations are based on the impressions thus gained. Since they have not been arrived at by asking identical questions in all 58 cases, it would be misleading to give the responses in percentages.

1. Sample

The sample selection was undertaken with the view of interviewing approximately 15 children in the two lower age groups (10 and 14 years) in both England and Germany, with a reasonable spread for sex, intelligence, and socio-economic status.

Contact to schools was obtained through the N.F.E.R. in London and the UNESCO-Institute in Hamburg. Both institutions encountered great difficulties in getting access to schools and were in fact unable to provide schools for the 14 year old age group. Thus the final selection was more the result of expediency than of design. All the older children interviewed were German and in schools contacted privately.

Altogether 58 children were interviewed, 38 in the age range 10-11, and 20 in the age range 14-15. Of the 38 younger children 21 were British and 17 German. In all, six schools were visited, 4 for the 10 year olds and 2 for the 14 year olds. All schools were co-educational state schools. One of the schools for the older children was a select school, a Mittelschule, catering for an intelligence range in between that of the grammar schools (Gymnasien) and that of the secondary modern schools (Volksschulen).

For the younger age group socio-economic status was controlled by selecting one school from a thoroughly working class area and the other from a socially mixed area in both London and Hamburg. This procedure proved to be impossible for the older age group because there was no choice of schools and schools tend to be larger and thus draw from a less uniform population anyway. In addition, the children were asked to state their father's occupation and the occupation they aspired for themselves. This combined information, when broken down into the fourfold Douglas SES classification, gave the following distribution:

Social Stratum Age	I	II	III	IV	N
10 year Old:					
Girls	-	6	10	3	19
Boys	1	8	9	1	19
14 year Old:					
Girls	-	5	4	1	10
Boys	1	1	6	2	10
N	2	20	29	7	58

In order to overcome the teachers' tendency to present only their brighter students and thus gain a better idea of the spread of intelligence, the following procedure was followed: at each school the headmaster was asked to select one form in the appropriate age range. The formmaster of this form was then instructed to select the best, the worst, and two average students for each sex. To corroborate the teacher's estimate of a student's position the students were asked themselves. In all but one school, student and teacher estimates coincided; in this instance the headmaster was teaching civics himself in the selected class and the formmaster presumably did not dare to present the dullest student. It was therefore arranged to see the four worst students from another form at the same school.

This, as well as the inclusion of two immigrant children at the London working class school, plus the occasional addition of 1 or 2 students in spare time inflated the original sample of 6 (schools) x 8 (selected students) by 10 students to 58 interviews.

It must be emphasized that the conclusions drawn are applicable to these 58 children only.

2. Suggestions for the Main Dimensions

As stated in the Preliminary Report, it seems advisable to split the broad area of Democratic Values into two big dimensions: that of democratic values proper and that of citizenship values. The reason for this partition are twofold:

a) it is possible to learn and internalize all kinds of democratic values, yet not be prepared to participate actively in the community;

b) likewise it is possible to be a good citizen actively participating in community affairs, but to do it in a non-democratic manner.

It follows from Document A that the aim of civic education in most countries is not just the transmission of democratic values, so that they are practiced at the individual level, but also their application to public life, in short that the intended end product is an active citizen holding democratic values. As the 'success' of civic education may vary in this respect, yet as this very point seems crucial in an international comparison, it was thought necessary to split these two aspects for heuristic purposes. To the reader of current political sociology it will seem unlikely that the two dimensions will vary independently, but their relationship has yet to be studied.

Both democratic and citizenship values should be subdivided into two big areas: one a mixture of cognitive and emotive elements, the other consisting of emotive elements only.

The mixed area 1 should deal with the actual knowledge about citizenship and democracy, where the questions actually contain the word "citizen" and "democracy". The impact of civic education will most easily show up in this area: how the concept of democracy will gradually change in the child's eyes from a mere word with a positive ring (10y.) to a maxim that has to be upheld regardlessly (14y.) and presumably to a more or less balanced view of democracy as a system of government among a variety of others (18y.); how the concept of citizenship will develop from that of a merely good person with some outside interests to that of a person with rights and responsibilities.

Area 1 should further deal with the evaluation of democracy and citizenship apart from the actual knowledge about the concepts. As the emotive relationship to the concepts seems to be established before any real information about them is acquired, this seems to be established before any real information about them is acquired, this seems particularly important. As for the democratic values questions should be asked as outlined by R.B. Farnen and D.B. German under the headings "Government Ideas" and "Democratic Ideology". As for citizenship values questions could be limited to the desirability of becoming a good citizen.

The emotive area 2 of both democratic and citizenship values will on the surface of things not deal with democracy and citizenship at all. It will never mention the concepts, but will include all kinds of items tapping the child's readiness to act in a 'democratic' or 'citizen-like' manner. The rationale behind the division into area 1 and 2 is simple:

a) the child may in time learn all the pros and cons of democracy, he may appreciate a democratic form of government, yet he may not at all apply its principles of tolerance, equality, etc., to his own individual behaviour;

b) similarly, a child may be quite sure of the attributes of a good citizen, yet he may not see himself as ever filling such a role.

Several advantages can be gained from this division. First, one can see the depth of civic education in that one can compare the explicit teaching (via child's information in area 1) with the degree of internalization the information has reached in area 2. Secondly, one will get at least an impression of earlier socializing agents and their impact on democratic and citizenship values, because area 2 for the youngest age group will give clues about earlier socialization experiences. And thirdly, given the results of area 2, one could form an idea about the necessity of civic education - assuming of course the overriding importance of the emotive aspect over the cognitive for the persistence of the system.

Area 2 of the democratic values should comprehend a continuum on equality-inequality, for instance the child's views on and behaviour toward favouritism, a continuum on tolerance-prejudice (this could include a measure of ethnocentrism, if so wished,) and a section on 'freedoms', like freedom of speech, assembly and demonstration, travel, choice of job, etc.

Area 2 of citizenship values should comprehend at least one measure expressing the child's feeling that his actions will matter within the political context, like 'political efficacy', 'subjective competence', and the 'confidence index'. It should further include a continuum on political apathy/participation, of obedience (acceptance, rigidity) /disobedience (rule-breaking, criticism) and of trust/cynicism in people and politicians. These subdivisions should all be in the form of continua, where the one end will signify a good citizen and a democrat and the other end the opposite.

This is not the place to speculate, but the comparisons of the ratios of democrats to good citizens should open an interesting field for cross-national studies like this one.

- - - - - - - - -

Apart from these two main dependent dimensions two independent variables should be included: politisization and relation to authority.

By politisization we mean the child's actual behaviour toward political objects: his reading, viewing, discussing of political topics, his listening to and awareness of other people's political views. His professed interest and his party allegiances belong here likewise.

Relation to authority is a somewhat ambiguous dimension, as it could be treated in different ways: it could be looked upon as a dependent or independent variable, indeed it could be included in area 2 of citizenship values on the continuum obedience-disobedience. However, since the child has to establish some relationship to authority long before his schooling, let alone civic education, starts, it is recommended here to treat it as an independent variable. Possibly it would pay to separate relation to non-public authority (parents, maybe teachers) and relation to public authority (policeman, judge, Prime Minister).

Lastly, two topics were found to be highly productive of opinions and speculation amongst the children: elections and laws. All kinds of curious misconceptions are rampant here. It would perhaps be interesting to include them as dimensions in the cognitive domain of civic education. In the affective domain they will of course be touched upon by individual items, but not be treated as separate dimensions.

To summarize the recommendations for the main dimensions to be studied, here is a schematic drawing of them:

Suggestions about methods

Suggestions will be only on the type of question thought to be most suitable for this kind of research.

i. Attitude statements should be kept to a minimum at least for the youngest age group. Of course, even the dullest 10 year old child can easily tick Yes - No, Agree - Disagree codes, but all the children will be sorely tempted to cheat, i.e. make ticks without understanding the item. German and English children admitted that they would have cheated in this way, if they had been left to their task on their own without an interviewer asking "why" afterwards. In fact, some of them tried to do it in the interviewer's presence. Ways to overcome this problem:

   a) add a category like "I do not understand the statement", or "I do not know word __________". However, this creates additional problems. On the one hand there is the temptation to tick the non-committal category for many other reasons than lack of understanding. Secondly, the younger children, who doubtlessly have to tick this category rather frequently, will get frustrated by it.

b) check the consistency of answers by comparing the scores on items formulated once positively, once negatively. As this method also helps to overcome response set, it might kill two birds with one stone. However, when it was used with the F-scale, didn't even adults have a tendency to give inconsistent answers?

c) again use positively and negatively formulated items and administer them as one forced-choice item. This turned out to be far too difficult for the children. Quite often, the statements were not seen as opposites and to reflect about two complicated sentences at the same time was quite a task even for the 14 year old children.

d) to include some lie-detectors in such a way that a Likert-type item will be checked by a closed question. This can of course only be done for a limited number of items, but it should produce a measure of the truthfulness of the responses. This would come closest to the interview situation, where the interviewer can follow up each answer.

ii. Wherever possible items should be formulated in such a way that the codes indicate frequency, degree, prevalence, etc. Children find it easy to answer questions of the "how often", "how much", "to what extent" type. Bronfenbrenner's description of parental behaviour is a perfect example. The Hess/Torney questions about authority figures illustrate the same point, although shorter codes would have been better still.

iii. Reading material should be kept to a minimum too. For some of the 10 year old children reading is still a task that takes its time. Although this is obvious, it is a great pity, for children enjoy lengthy story-type questions (morale boosters) and in order to get at democratic behaviour, quite detailed descriptions of situations are necessary. So a happy medium will have to be struck on the question of reading material.

iv. Visual material is a great help for the young children, but not essential. Again as a morale booster an occasional picture helps, it is also useful if the telling of a lengthy story is thus avoided and the child can grasp the situation at a glance as in the Rule-breaking Test, or where pictures of national symbols are used, which are familiar to the children by sight, but not by name. Yet in this last instance the beauty of pictorial material, the avoidance of problems of translation, is lost. If questions about social classes are intended, a schematic drawing of a population pyramid with layers proved a great success. However, in general things can and should be left to the child's imagination. Pictures of soldiers, teachers or policemen are superfluous, if anything they raise more problems of the "our teacher does not look like this" type.

v. Closed questions are better than open-ended ones. There are obviously advantages and disavantages to both. Open-ended questions are much better for pilot work, but they impose the tremendous difficulty of writing sentences on the child. They may have their place in the questionnaire for the older age groups, but for the 10 year old children they should be excluded, if only for the time it takes the child to write down his answer. The closed question suggests answers to the child which might never have entered his mind, and it limits originality and variety. On the other hand, it facilitates the link-up of hypotheses and items, the coding and the analysis, so that the balance the odds are in favour of closed questions.

4. Differences of Age and Intelligence

To sum up: differences of intelligence were more striking than those of age with "interest in politics" a powerful intervening variable.

Among the 10 year old children there were some who did not compare unfavourably with some of the older children, yet there were others who could not talk about anything further removed than their home or school. The only real and persistent difference between the intelligent 10 year old and the dull 14 year old was that the older children had a larger passive vocabulary to deal with political matters. Although they could not express their point of view any more eloquently, they could understand many more attitude statements, which will of course be a great help. On the other hand, this increased vocabulary can lead to some real confusion. For example, the older (dull) children had been taught to differentiate between democracy and dictatorship, and this lesson seems to have been brought home as something really vital. So when talking about the state and the government, the following descriptions were given: State - that's where everybody has a say, Government - that's where one man rules.

On the whole the 14 year olds had of course far more information on topical political questions, and the teaching of history and civics had enlarged their knowledge of concepts. Yet with them the main differentiating factor was not so much intelligence any more, but whether they were interested in politics or not. The interested youngster was far more advanced; when asked to state his views on communism for instance, one boy replied: "Which shall I deal with first, the Chinese, the Russian, or the Israeli communism?" As a result, the two most 'advanced' children were a boy at the select school in danger of having to repeat his year and a girl at the top of her form in the non-select school. Both of them could make intelligent responses to the Questionnaire 3.C. for students taking their A-levels (Primaner).

5. Sources of Attitudes

Children obtain these values from many different sources apart from their home and their school. The sources can be as varied, as can possibly be imagined. Visitors to the home, relatives, neighbours, mass media, recent political events and legislation, etc., etc. To cite some examples: "in a book in the office of my father's boss, there was something about these men who tried to kill Hitler", "my auntie lives near some black people and you want to hear her talk about their fights", "my mother's girl friend's father was a NAZI, so she always had ham on her sandwiches", "I like listening to radio Moscow and radio London". It does seem from this that strange and definite sources are well remembered, whereas the more usual sources, from which information flows more or less continuously, are not mentioned. Questions about the relative importance of home, school, friends, etc., in transmitting attitudes would therefore produce rather misleading results. Although knowledge on other sources of attitudes is necessary in order to delineate the effect of civic education, which is but one source, and although it would be interesting from a therapeutic point of view, it is felt that this task is better not attempted.

6. Observations on Children's way of Thinking

It has to be pointed out that children seem to follow some peculiar patterns in their perception and understanding of the adult political world. Some of these patterns have influenced our choice of dimensions, others can be used as hypotheses, yet others may be useful in the selection of items and precodes.

i. Children tend to organize everything into good or bad without really knowing anything about the object so classified. This was tried out with the concepts of 'democracy' and 'communism'. Overwhelmingly the English and German 10 year olds thought democracy to be something good and communism something bad, although hardly anyone of them knew anything else but the mere word. This suggests that emotive learning precedes cognitive learning. If this is so, one can hypothesize that, depending on the direction of the emotions civic education will either be reduced to reinforcing already existing notions or it will have a hard time trying to change these early categorizations. It also suggests a rather terrifyingly easy procedure to indoctrinate children. - The older children, while maintaining the emotional loading of the concepts, had something to say in justification of their views.

ii. Children group people in rather stereotyped big black and white categories. For England the categories were the dichotomy 'nice-horrible', which was used by the English as well as the two immigrant children. If people were described in any other terms, such as 'fighting-non-fighting', or as 'giving you things or not', it could always be reduced to nice or horrible. In general, any group had nice and horrible people attributed to it. - For Germany the stereotype was again 'nice', but this time with all its variations: 'not so nice, somewhat nice, quite nice',etc. Again, there are always nice and not so nice people in any group. So in a way one could speak of tolerance of diversity, the problem is that the diversity is not consciously perceived. By age 14 this grouping into stereotyped categories had more or less disappeared, differences between people apart from nice and horrible had become apparent.

iii. Children work with the concept of reciprocity, or the belief that nothing is done on its own merit but for hard and tangible reasons. For instance: "if we let immigrants into our country, their countries have to buy our exports; foreigners should be treated well, because English people expect to be treated well abroad". This tendency seems in fact to increase with age, and it is by way of reciprocity that children first come to grips with the idea of interdependence and involvement. Starting with interdependence between nations in terms of economics and trade, they proceed to interdependence between people within one nation: "farmers abuse businessmen, businessmen abuse farmers, but none can do without the other".

iv. Children find it hard to distinguish between the 'IS' and the 'OUGHT TO BE'. Whereas the majority of the 10 year old children would never have perceived a discrepancy between the actual and the ideal, several of the older children made this point of their own accord, when asked to describe a figure of authority. For the young ones the way something is done is the right and proper way to do it. It is this that seemed to speak against the inclusion of a dimension like 'legitimacy'. The most intelligent English boy going by IQ scores, for instance, could simply not imagine an unsuccessful government: "it just would not be the government". -This phenomenon could be used in two ways. First as a tool by asking the children to describe an ideal and an actual authority figure. Secondly, as an indicator of maturity, if such a tool produced differences for the actual and the ideal.

v. The young child is not yet spontaneously critical of political matters. This may not be surprising after what was said under point iv. and in the fact of the child's rather scanty information, yet want of information does not stop him from categorizing into good or bad. The little spontaneous criticism there was mostly came from English children. Their target was the Prime Minister: "him and all his laws ", he isn't nice when he puts up the prices". The only critical remarks from German children concerned the President. With increasing age the targets for criticism become more varied and the expression of it freer. This readiness to criticize has been included as one end of a continuum under citizenship values.

vi. In their thinking of every aspect of governmental functions children tend to personalize and concretize. They either think of the incumbents of public offices as doing things for purely personal reasons or attribute events to personalities only. For instance, Johnson took over Kennedy's job, because he was his brother; the wall in Berlin was built by that chap in the goat's beard, Ulbricht. Institutions like the state and the government are seen as 'big houses' and they in turn are personalized, e.g., if the government made a mistake, the landlord of the Houses of Parliament would throw them out. In fact this 'parliamentary landlord' figured prominently in both countries. -This tendency diminishes with age and it could therefore be used as an indicator for maturity. Where it persists, it occurs in a slightly more sophisticated way, e.g., the child would consider the possibility that his country was in the wrong, if he had relatives on the opposing side. The concretization of institutions seems to disappear more slowly, as there were still some 14 year old children, who thought of the state as an office in Bonn.

vii. The 10 year old child holds his views without any attempt to justify them. Whenever a child had pronounced a clear-cut opinion on an issue, he would be at a loss to answer the Why-question. These opinions were isolated ideas which the child must have picked up somewhere without a coherent ideology behind them, rather like the child's party allegiance and his ignorance of party ideology. -By age 14 this had changed a lot, reasons for opinions were volunteered and attempts were made to link up responses with earlier ones. In fact some children could give a running commentary when answering one of the scales.

7. Observations on Children's Perception of political Objects and Problems

Children relate themselves in fairly definite ways to the political objects surrounding them, from filling them with distinct meanings for themselves to hardly noticing them at all. These patterns are listed, as they may suggest some hypotheses.

i. Children are not aware of the community as such. The preliminary report stated that they cannot see social problems, cannot imagine how things could be different and do not perceive any oragnizing principle behind the everyday public objects they come into contact with, in short that everything is taken for granted. This has got to be qualified. For all the children see themselves as part of a community: their country and their town. In Germany the knowledge of being a German was usually accompanied by a lot of confused ideas about the division of their country (into north and south, into Berlin and Germany, into the DDR and a Germany that had been amalgamated with Belgium and Switzerland). Social problems had entered their world as far as children were involved, e.g., starving children, children who had been kidnapped, and as far as they have come into direct contact with them, like a chronically sick father or an old age pensioner living next door. Usually though such problems are far removed. -For the older children social problems had made their appearance, but still only sporadically: government should provide work for invalids, it should give money to the sick and provide a holiday for everybody in need, or stop arms purchases in favour of supporting the poor.

ii. Social inequality has hardly been noticed by the young child. To free questions about differences between people - apart from the already mentioned nice-horrible distinction - children think of differences between boys and girls, between children and adults, and only after after some prompting between coloured and white, rich and poor people. Comments about these differences show that at least the last one is not seen as an organizing principle of society: "poverty is only temporary", "it's their own fault, they must have been drinking", "they'd get over it, if they would only work", "the Queen is rich, but I am half rich". Questions about social classes are marred by terminological problems: in England class was taken to mean a school class, in Germany a worker's shift (Arbeiterschicht) or a layer of some sort. When the term was understood, the dividing lines between the child's social classes are quite original: "working class are those who work and non-working class are those who are paid by the government and all the workers' wives", "I am lower class now, because I am still a child, but I'll be middle class as soon as I am grown up", "my Dad is industry class, I don't really know what my mother is, she works in a children's home". If a schematic drawing of a pyramid with layers labelled 'rich' and 'poor' is used, the children found it easier to work with the term social class.

To the older children social stratification had become very apparent indeed, but strangely enough its existence was denied rather often. The first response was usually that classes are a thing of the past, that people were equal today, that there were no more slaves and castes, and that the lower class was so small, as not to deserve any attention - the workers were now all in the middle class. Nonetheless inequality before the law was mentioned astonishingly frequently, (the rich being able to bribe the judge, etc.) and the unfair advantages the employers had over their workers were bemoaned.

iii. Children are unaware of the existence of adult conflict. Conflict is seen only on the individual level, like boys fighting in the playground and adults arguing across the fence. They cannot conceive of clashes of interest between groups, such as unions and employers, religious groups or conflicting business groups. The world of the grown-ups is all harmony. Consequently they have no idea that much of governmental work is concerned with the regulation of competing interests. Children do not think of adults as members of possibly competing groups, everything is perceived on an individual level.

For instance 'strikes': the English children had heard about them more vividly than the German ones, but these strikes weren't a problem in their country, maybe in the United States. Two German boys reckoned they were fights among workers, with the employers figuring as peace makers.

The exception to this general picture of harmony is war and oddly enough for the German children an uprising. Wars are fought about such tangible things as rivers and fertile land, they are caused by quarrels among politicians, but they do not really happen anywhere near the child's own country - this may be accounted for by the topicality of the Arab/Israeli war. The idea of an uprising and of civil disorder among the German children appeared in various contexts: public criticism of the government would cause an uprising;if children were not loyal to their country, it could be the start to an uprising; Communists had something to do with uprisings; small religious minorities were a danger, as they might disturb civil harmony; if there was only one party, people would revolt or perish. One little girl only read the first page of the local newspaper to see whether there had been an uprising or not. -By contrast the word uprising was never even mentioned by the English children and only occurred once among the older German children.

Perception of conflict had become more varied for the older children. For them conflict could be caused by religious differences, competition in industry, and excessive complaints. Clashes of interest within one country that were a bit out of the ordinary were: the people against the government, enlisted men against the army, and students against the police.

iv. Children have no distinct ideas about settling conflict. With conflict at the individual level, children would either report it to someone with more authority or they would simply join in and/or retaliate in the same fashion. With more distant incidents of conflict, like wars, their suggestions about solutions are limited to verbal recommendations like "they should stop", "they should make peace", etc. Compromises, negotiations or majority decisions to settle conflict are not very salient. Occasionally a child mentioned that discussion might help, "if they discuss well". No national differences on this score.

4 years later the picture has not changed significantly. More emphasis is put on finding out the guilty party, the one who started. But the tactics that should be applied after that are still reliance on authority and force (the Bundestag should listen to both sides first and then intervene with force).

v. Children have trust in the leadership of well-known politicians. They attribute charismatic qualities to some of them. The Queen for instance could and should make a law to stop all wars; if Adenauer, Kennedy and de Gaulle were still alive, there would not have been a war between Israel and the Arabs just as there were no wars during their life time. Or: the Prime Minister could satisfy everybody, if he only wanted to and if the country was split 50-50 on an issue, he would be clever enough to decide it. This trust was not found with any of the older children. It raises an interesting question in connection with the lack of criticism (see 6.v.), for if a child did criticize spontaneously, it was criticizing the Prime Minister. One possible explanation: these patterns are sequential. To start with trust, then the first attempts at criticism, then the disappearance of trust.

vi. Another explanation is possible: Children make a distinction between the head of state and the government, however vaguely. Whenever this distinction is made, all the trust and affection is placed in the head of state leaving the prime minister free to be criticized. If the distinction is not made, the trust is extended to the Prime Minister and his government. Among the German children the distinction was more confused and twofold in nature. The President and the Chancellor are seen in an hierarchical relation but with distinct areas of influence. Apart from this a difference is seen between the state and the government. This time the state as the giver of pensions, sick pay etc., is invested with all the affection and trust. A caution must be added: the term 'state' was misunderstood by almost half the children, because of the German colloquialism "Vater Staat". It was actually seen as a person, as an organization for fathers and as a federation (Vereinigte Staaten). -With two of the duller 14 year old children this functional differentiation of state and government persisted (state: home affairs, government: foreign affairs), to the others the concepts had become clear and the offices of President and Chancellor were more accurately defined.

vii. As fits the general status of a child, he does not see any justification for civil disobedience of any sort. In the child's eyes nobody has got a right to quarrel with a policeman whatever the circumstances, and no child should disobey a teacher. To break a law is bad whatever the reason and only crooks do it. - Age differences are pronounced in this area, as at 14 quite a few could conceive of situations where civil disobedience might be advisable. Student demonstrations, the formation of groups to fight specific issues, and girls who do not always do as told, were condoned by them.

viii. Children have very little time perspective and no sense of history. They do not understand the concept of progress and are quite unaware of social change. For instance, laws last about 3000 years, if made by Parliament, and about a million years, if made by God. This links up with the way they take the world, as they know it, for granted and assume it always to have been the same, permanent and unmodifiable. - Though the 14 year old children had acquired more historical knowledge, the ideas of social change and progress were still quite hard nuts to crack.

ix. Children differed by nationality in the way they saw the relationship between government and economy. The English youngsters considered the determination of prices and intervention in the economy to be the main function of the government, though not quite in these terms. The Germans, by contrast, never thought that the government might have anything to do with prices. Yet by age 14 the German children assumed a dependent relationship to exist between the economy and government, more specifically a democratic government. They all took it for granted that the government was in a shaky position, if the economy was not doing well: people would lose all interest in the government, poor people would always want more money from it without caring what form the government would take. The present crisis was seen as a test case.

x. Children differ by age in the importance they attach to the image of their country. This could of course only be observed with the German children, as no older English children were interviewed. The younger German children never seemed to have given their country's image a thought, all the more surprising was the preoccupation of the older ones with this question.

Not only did they realistically perceive the reservation in other countries toward Germany, there were also estimates that up to 40% of the population, mainly the young people, did not feel any pride in being German; interest in foreign affairs was justified with the need to make friends; one girl manifested what could only be called a persecution complex when she lamented that whatever Germany would do or did, it would be held against her.

xi. The older German children were full of democratic slogans. These were slogans that never turned up with the younger ones, like "we all are equal", "equality of rights in a democracy", "communists are also human", "one has to learn to respect other people's opinion", etc. They also showed eloquent disapproval of dictatorship and immediately made the association democracy-dictatorship. This turned out to be the influence of their civics course, in their view. It would have been interesting to see whether these slogans had been internalized, in other words to see whether they would have scored as 'democrats' on the measures proposed for area 2 of the 'democratic values'.

List of Stimulus Material

Pictorial Material: Children's Playgroup with teacher, Nuns giving food to old people and tramps, UN building, UN general assembly, U Thant, Student sit down demonstration, Egyptians demonstrating against Intrepid, Soldiers in action, Sequence of Hitler giving a speech, Numerous pictures of English, German and internationally known politicians, Pope, the Queen, National Symbols: the flag, Houses of Parliament, Nelson's column, No. 10 Downing St.

Scales: Belief in the Bill of Rights (Remmers), Trust in People (Rosenberg), Conservatism/Radicalism (Centers), Manifest anti-Communism (Centers, Harvey), P.E.C. (Adorno et al), Latent Marxism (Remmers), Political Extremism (Harvey), Dogmatism (Rokeach), F-Scale (Adorno et al.), R.-T. factors (Eyseck), Patroitism (Harvey), Anomy, (Srole), Sense of Political Efficacy, Sense of Citizen Duty, Subjective Competence (Almond/Verby), Confidence (Oppenheim), Parental Behaviour (Bronfenbrenner), Rigidity (Leach), Political Apathy (Rosenberg, Harvey).

Selected questions from:

The Development of basic attitudes and values toward Government and Citizenship during the Elementary school years, by R.D. Hess, J.V. Torney; "Zur Wirksamkeit politischer Bildung", Institut fur sozialforschung an der Goethe Universitat, Frankfurt;
Hochschule fur Internationale Padagogische Forschung, Fragebogen fur Primaner;

Varied: Tendency to generalize (Leach),
Rule-breaking Test (Leach),
Moral Judgement Test (Leach),
Semantic Differentials of Politicians, and Teachers.

APPENDIX I

SPECIFIC SUBJECT MATTER HYPOTHESES AND QUESTIONS

SPECIFIC SUBJECT MATTER HYPOTHESES AND QUESTIONS

This is a list of the hypotheses and questions suggested by the subject matter committees.

i. Science

1. To what extent are the newer science teaching methods based upon investigation and student enquiry achieving better results than traditional methods?

2. Are any differences observed more marked in achieving some objectives than others?

3. To what extent is the success of science teaching, and particularly the newer methods of science teaching, dependent upon

a. the material conditions under which science is taught, e.g. provision of laboratories and apparatus

b. the time allocated to science and the fraction of it spent in practical work in the school timetable,

c. the provision of laboratory assistance,

d. the training teachers have received and the facilities for refresher courses?

4. Is a more adventurous attitude towards future careers in science, as indicated by a wider range of desired careers, favoured by the newer teaching methods?

5. What differences can be detected between the sexes in their attitudes to, and their achievements in:

a. science as a whole,
b. the physical sciences and the life sciences.

6. To what extent are any of these differences related to:

a. social variables,
b. home background, including conditions and interests,
c. school administration and organization
d. teacher variables?

7. To what extent is achievement in, and attitudes towards, science related to:

a. the level and rate of advance of technological development in the countries concerned.
b. rural and urban environments?

8. To what extent is achievement in, and attitudes towards, science related to school organization variables such as comprehensive and selective systems of education?

9. The highest levels of science achievement will be attained by a small percentage of full-time students in countries in which a smaller proportion of the relevant age group is still attending school.

10. Where the lower mental process score is held constant, performance on the higher mental process score will be directly related to:

a. the quantity and quality of the training of the teachers.

11. Where the knowledge and skill achievement in science is held constant, performance on the more complex objectives will be directly related to:

b. teachers ratings of the student's opportunity to learn items involving more complex objectives.

12. Achievement on basic science skills and basic content is unrelated to socio-economic status of students (while achievement on the more complex objectives is directly related to socio-economic status).

13. The role of women in a society as indicated by occupational freedom and educational opportunities will determine the differences in the interests, attitudes and achievements in science of boys vs. girls.

a. In all countries, there will be no differences in overall science achievement of 13-year-old boys and girls but there will be slight differences favouring the girls on highly verbal problems and slight differences favouring the boys on computational problems. (Hold constant the level of science instruction).

b. the following will vary from no difference to marked difference depending on the role differentiation of women in the countries.

i. girls will differ from boys in interest in science;
ii. girls will differ from boys in plans to take further science;
iii. girls will differ from boys in attitudes toward science;
iv. girls will differ from boys in view about who can learn science;
v. girls will differ from boys in science achievement;
vi. girls will differ from boys in the higher and lower mental process scores.

14. The above differences between boys and girls will be greatest when the students are in single-sex schools; the differences wil be least when they are in co-educational schools.

15. The above differences between boys and girls will be greatest for students whose parents have completed only elementary education while the differences will be least for students whose parents have completed secondary or more education.

16. Indices of the nations' urbanization, geographical and occupational mobility, and technical-industrial development of the countries, will be related to:

a. science achievement at terminal stage, but not at 14-year-old level.

b. interest in science and desire to take more science

c. attitudes toward the role of science in the society

d. description of science teaching and learning

e. view of science as a process

f. views on who learns science

g. attitude toward school learning

h. performance on higher mental processes, but not on lower processes.

17. Students who plan (or desire) to enter scientific and technical occupations will differ from students who plan (or desire) to enter humanistic and non-technical occupations in:

a. interest in science

b. wish to take more science

c. attitudes toward the place of science in society

d. attitudes toward a science as a process

e. attitudes about the difficulties of learning science

f. total science score

g. higher mental processes

h. lower mental processes.

ii. Reading Comprehension

1. Internal structure of the test battery:

a. What are the relationships between the different tests of reading comprehension, in the different countries and how do these relationships change according to the level of education of the target populations?

b. What are the relationships between reading comprehension, rate of reading, vocabulary, adjustment of rate of reading, in the different countries; how do they change according to the level of education of the target populations?

2. Educational practices and background:

The level of reading comprehension may be influenced by:

a. methods of instruction at the beginning of reading (i.e. analytic, synthetic, mixed)

b. pre-school instruction in reading

c. importance of reading in the curriculum (definite reading lessons)

d. formal training in grammar, in vocabulary

e. kinds of reading assignments if any (literary appreciation, extensive reading, memorizing poems or texts, etc.)

f. number of teachers the child has each week.

g. type and size of school

h. type of studies (i.e., general, technical.....)

i. specific subjects studied

j. grade

k. advance or retardation in studies

l. the grade at which a second or foreign language is introduced.

3. It may also be influenced by the availability of reading material:

a. in the school (does the child have text books for different subjects, is there a school library?)

b. in the home (books, newspapers, magazines)

c. in the community (can books, newspapers, magazines be bought or borrowed?)

4. The environment outside the school may have an influence on reading comprehension:

a. Occupation and social standing of the parent(s)

b. Level of instruction of the parent(s)

c. Amount of exposure to mass media (radio, TV, films)

d. Is the language of instruction the child's first language?

e. Is the child monolingual?

5. The interests of the child and his attitudes may have an influence on reading comprehension:

a. Preference for some of the subject matters (by direct ranking, by ranking of the different texts of the test, etc.)

b. Interests in leisure activities

c. Attitudes toward reading (child's idea of the aims of reading, why it is important to be able to read, etc.)

d. Ambitions and aspirations of the child

e. Attitudes of the child toward school (does he regard it as a compulsion)

6. Study and reading habits of the child may influence his reading comprehension (e.g. finger tracing, vestigial oral reading, etc)

7. There probably is a relationship between the child's achievement in different subjects and/or overal-achievement and his level of reading comprehension.

8. Physical defects might impair the child's reading. (e.g. eye sight, hearing, speech defects)

iii. French as a foreign language

1. (i) Variation in achievement profile will reflect variation in teaching methods and objectives: children taught French by audio-visual methods will reach a higher level of achievement in speaking and listening skills; children taught French by grammar-translation methods will reach a higher level of achievement in reading and writing skills.

(ii) This variation will not persist into the pre-university grade (if length of exposure to French is held constant)

2. Level of achievement in French will be related to length of exposure to French rather than to chronological age or grade level.

3. Attitudinal factors (in both home and school environment) will affect level of achievement in French.

4. Socio-economic factors (parental education, socio-economic skills, etc.) will be related to child's level of achievement in French.

5. Girls will reach a higher level of achievement in French than will boys: this difference will be accentuated in single-sex schools.

6. Size and composition of French class will be related to achievement in French.

7. Selection procedures will be related to level of achievement in French: children in streamed schools will achieve a higher mean score than will those in comprehensive systems; the latter will produce a higher total "yield" but a lower average score.

8. Children's level of achievement in French will be related to the teacher's training, qualifications, and teaching experience.

9. Variation in economic support for education will be related to variation in level of achievement in French.

10. Children's level of achievement in French will be related to parental aspirations and attitudes.

11. Level of achievement in French will be related to opportunities for contact with French speaking groups.

12. Level of achievement in French will be related to the "linguistic distance" between French and the mother tongue.

13. Measures of verbal ability in the child's mother tongue will correlate highly with measures of achievement in French at the advanced levels; in the early stages of learning, there will be a much lower correlation.

iv. English as a foreign language

1. The greater the amount of assistance provided by the country in teaching English (in terms of guidance, supervision, facilities and equipment), the higher the educational achievement in English will be.

2. The degree of linguistic complexity of the country (number of languages or distinct dialects spoken) will affect the level of achievement in English, especially in the early stage of learning English.

3. The linguistic distance between the country's (native) language and English (as established by expert linguists) will be negatively correlated with the level of general achievement in English.

4. The country's need for English as a medium of communication in international trade (including tourism), science and technology will be positively correlated to (i) the achievement in English and (ii) the amount of assistance provided by the country - see also Hyp. 1.

5. The quality of teacher training will be positively correlated with the achievement in English.

6. Schools having special English teachers will perform better than schools having teachers teaching other subjects besides English.

7. The attitude of the teacher will be positively correlated with the students' achievement in English.

8. The effect of size of class on the achievement in English will vary with the different sub-tests.

9. The method of teaching will affect the level of achievement in English.

10. The linguistic background of the student will affect his level of achievement in English especially in the early stage.

11. Age at which the student starts English will be negatively correlated with his achievement in English - holding constant the number of years of English training, those students who started at a younger age will do better than those who started later.

12. There will be an optimum range in the degree of concentration in learning English (or intensity in terms of the number of hours of English classes per year), and the relationship between degree of concentration and level of achievement will be curvilinear.

13. There will be a positive correlation between the socio-economic status of the parent and the student's achievement in English. The correlation, however, will be higher than that between socio-economic status and achievement in mathematics.

14. The general ability of the student, in terms of (i) grade point average, and (ii) I.Q. will be positively correlated with his achievement in English. The correlation, however, will be lower than that between general ability and achievement in mathematics.

15. The attitude of the student toward the English language and toward the culture associated with the language will be positively correlated with his level of achievement in English.

16. Attainment in particular aspects of English - reading, writing, understanding and speaking - will vary directly with the kind of method employed in teaching - grammar/translation, or audio-visual. (c.f. Hyp. 1)

17. Attainment among individual students will be positively related to the degree of specialized selection employed in forming a class.

v. Literature

1. Characteristic patterns of expressed response will be closely related to knowledge about and attitudes toward literature.

2. The various outcomes of literary education are related to, but not identical with each other.

3. (i) International differences in the stated aims of literary education are related to the place of literature in a nation's culture and practices of literary education.

(ii) International differences in the outcomes of literary education are more closely related to the practices of literary education than to the official statement of its aims.

4. Within and between nations, outcomes of literary education will vary in relation to social, cultural and individual characteristics of students.

5. Internationally, the differences between elite and popular culture as represented in the schools is related to the place of each culture in the society as a whole.

vi. Civic Education

1. Student knowledge of governmental affairs is higher when civic education is given as a separate subject than when it is an overall aspect of a number of subjects.

2. The students' interest in participating in social activities is positively related to their interest in governmental affairs.

3. Students' participation in extra-school activities (clubs, games, etc.) is related to their perception of authority relationships within the family.

4. The more a curriculum in civic education stresses a functional approach (as opposed to a structural approach) the greater will be the variability of the knowledge-profile pattern.

5. The variability of knowledge of international relations is positively related to the degree a curriculum in civic education stresses a structural approach

6. The student's interest in the subject civics is higher when instruction is given in short, concentrated courses, than when instruction is spread out over a whole year (or more).

7. Student differences between age 14 (±1) and age 17 (±1) in attitudes towards participation in social activities is positively related to the type of school organization which promotes such activities.

8. There is a negative correlation between the students' knowledge of international organizations and their thinking in stereotyped terms.

APPENDIX J

THE ANALYSIS AND PRESENTATION OF THE IEA EVIDENCE

THE ANALYSIS AND PRESENTATION OF THE IEA EVIDENCE

Our final object is to write a clear and agreed report in which our evidence is presented and analysed, so that we can draw some conclusions and raise some questions. This paper considers some of the problems that we must overcome on the way to our goal.

I. DESCRIPTION

1. How Should we Present the Weights.

Our evidence will be derived from samples, but our interest lies in the populations that these samples represent. Our samples should be probability samples, which means that every pupil, teacher and school in a population should have a chance, which we can specify, of appearing in the sample. The chances are specified by the overall sampling fractions, and the reciprocals of the sampling fractions, which are called the raising factors, tell us how many pupils/teachers/schools in the population are represented by a pupil/teacher/school in the sample. In the analyses either the raising factors themselves, or any set of numbers proportional to them, can be used as the weights. The constant of proportionality makes no difference. But for the presentation we must decide what constant would be least likely to mislead the reader. Among the possible choices two stand out. Thus in the case of pupils we could either choose the constant so that the sum of the weights was equal to the actual number of pupils in the sample, or we could choose it so that the sum of the weights was equal to the number of pupils in the simple equivalent sample. The ratio of these two is the design effect, and its square root is the number by which simple random sampling estimates of the standard errors should be multiplied to give the estimates of the actual complex standard errors. If it could be assumed that the design effect were well understood there would be no problem. But this is not yet the case, as was clear from some of the comments on our first report, and on the whole the safer course seems to be to make the weighted sum equal to the actual sum, and to print this together with the design effect. The decision on this point should be taken before the evidence is put into the computer, although the design effects will only be known after a certain amount of analysis has been done.

2. The Component Analysis.

At an early stage the component analysis between the country means of the part scores be carried out, to see whether the part scores need to be re-weighted. If the component weights do not differ appreciably from the assigned weights no re-weighting will be needed. But if re-weighting is required it should clearly be done at the earliest possible stage.

3. Analyses Over All and Between and Within Schools.

Within each country we can make three kinds of analysis. In the first place there is the kind which for want of a better name is called over all. In this the measures for each pupil are used, disregarding (except in the calculation of the design effects) the fact that each pupil is a member of some school. Secondly we have analyses between schools; for which the measures for pupils in the same school are aggregated, so that differences within the same school disappear from the analysis. Finally there are analyses within schools, for which the school mean is subtracted from each pupil measure, so that differences between schools disappear. Analyses between and within schools are complementary to the analyses over all. The first stresses the differences between schools and their neighborhoods, by eliminating the differences between pupils in the same school. The second stresses the differences between pupils in the same school, by eliminating the differences between schools.

4. Analyses Between Countries.

In principle we could also put together the evidence from different countries by analyses over all, between countries and within countries. Whether this would be useful depends upon the relative extent of the common features and the differences between countries. If the differences greatly preponderated a more piecemeal procedure might be preferable. In any case it seems clear that in putting together the evidence from different countries each country should be given the same weight. There is little to be said for making the country weights proportional to population, and nothing to be said for making them proportional to the size of the sample. On the other hand it seems better to add covariance rather than correlation matrices.

5. How much Detail?

How much detail should be printed in the computer output? There are two aspects to this question. On the one hand it is plainly necessary that members should have the fullest possible access to the evidence, so that the agreed report can be based on the combined wisdom of the whole body. On the other hand the report itself must not be swamped by detail. This means that the output for circulation to members should be more extensive than the output included in the final report. But how much more extensive?

6. Univariate Statistics.

The first requisite will be the information about the measures separately - the univariate statistics. A member will need to know the name of each variable, and its class (e.g., pupil or school), its type (e.g., scaled or unscaled) and its sign. It would be convenient to list the classes separately, and within each class to put the scaled variables in a first and the unscaled in a second block. Unless the sign is indicated doubt may arise (e.g., for sex of pupil). The mean and standard deviation could conveniently be given in the same list, and a following list could give the percentage frequency distribution and the weighted N. For unscaled variables the second list could give the proportions in the previous categories, A, B, C, etc., while the standard deviations in the first list would be for the dichotomies A versus not A, etc. These are needed for interpretation if the regression coefficients are in standard form. Thus if boys count high in the variable Sex of pupil a regression coefficient of 0.1 means that boys score $0.1/\sqrt{pq}$ more than girls, where p is the proportion of boys and the scores are in standard form. In the general unscaled case this applies to all excess scores above that of the class taken as standard, so that it is convenient to have the standard deviations $\sqrt{pq}$ tabulated. A dichotomous variable has only two values, but in standard form they are not (0,1) but $(0,1/\sqrt{pq})$.

6.1 These are the univariate statistics for the analyses over all, and those for the analyses between schools and within schools could be displayed in the same way. Alternatively, and perhaps preferably, they could be indicated by merely adding, to the table of student variables, a column giving the proportion of the variance that lay between schools. For the school variables this proportion is of course unity, and so it is for the teacher variables, since these are only to be considered when averaged over schools.

7. Bivariate Statistics.

The next step is to obtain the bivariate statistics, or in other words the correlation matrices. These will be in three forms, namely, over all, between schools and within schools. Between schools the number of degrees of freedom is one less than the number of schools, and within schools it is the number of pupils less the number of schools, and in both cases s.r.s. formulae will apply for the standard errors. In the over all case the stability of the estimates can be obtained by replication, in the way indicated on page 62 of Bulletin No. II.

7.1 These correlation matrices will contain a great deal of the information about the relations between the variables, and will be the foundation for the subsequent regression analyses. A scrutiny will show which are the largest correlations and where these large correlations are of particular interest they could be displayed either as scatter diagrams or in tabular form. Display in these forms is illuminative and suggestive. On the other hand it needs so much space that it can only be applied to a fairly small number of carefully chosen relations.

8. Interaction.

The next step is to decide whether any, and if so what, product terms should be included in the regressions. The point here is that if there is an interaction between two variables it can only emerge from the regression if the corresponding product term is included. We know from past experience that most interactions are quite small, particularly when unscaled variables are used in cases where the assumption of linearity seems likely to be false. We cannot put all the product terms into the regression equation, because there are far too many of them; twenty variables give 190 product terms. Fortunately there has recently been developed a complementary mode of analysis whose chief merit is that it can reveal large interactions when they exist. This method proceeds by making successive splits of the total sample, which is finally resolved into a number of clusters such that each cluster is defined by ranges of values of certain of the predictor variables and contains pupils with a much smaller variation on the criterion than is found in the whole sample. The process begins by finding that predictor, and the split on that predictor, which divides the sample into two clusters such that the criterion variation between them is a maximum, and consequently the variation within them a minimum. It then takes one of these clusters and splits it again in the same way, and so on until an unsplitable cluster is reached - a twig, so to speak, at the end of a branch - after which it returns to reduce the other branches to twigs. A cluster becomes a twig when it is either too small or too homogeneous to be split further, or when none of the predictors will split it. In principle this process of resolving the whole sample into a set of most homogeneous clusters is much simpler than multiple regression, and it does not impose the same restrictions or assumptions on the variables. But it is also plain that in practice it is extremely complicated if there are more than two or three variables, and it has only recently been computerized through the work of Sonquist and Morgan at the Survey Research Center at Michigan. In their monograph (No. 35) these authors describe the program and also give the results when it was applied to nine survey examples previously analyzed in other ways, to illustrate its working. These examples show that while the analysis by successive splits is generally in broad agreement with the analysis by regression it does sometimes reveal the existence of large interactions which quite alter the interpretation. The test for the existence of interaction is lack in symmetry in the tree. If there is little interaction the tree will be symmetrical. If there is much it will be lopsided, with branches of different lengths. An attractive feature of the output is that each final group can be described completely, in terms of the predictor variables, by following the branch that leads to it. It may not be unduly sanguine to think that this may give us some new insights, apart from the question of discovering interactions.

8.1 This form of analysis is described at somewhat greater length in IEA/B/40 with a discussion of some of the problems of applying it. We should gain some extremely useful experience with it on the dry run, and it will also be applied to the second Plowden survey before the main I.E.A. survey takes place.

9. The Regression Analyses.

After the analysis by successive splits has shown what interaction terms, if any, are needed the regression analyses can be carried out. These will give the alternative description of the evidence. Putting the two descriptions side by side will be like looking at the photographs of the same landscape in different lights. Each may reveal features hidden in the other.

9.1 So far we have been concerned with the description of our evidence, in terms of the relations between the variables. These relations are exceedingly complicated, and economical and compact descriptions of them are needed before we can undertake the second and more difficult part of our task, which is to infer possible chains of causation. In this part of the task we must be guided not only by the immediate evidence which we have summarized, but also by our preconceptions, which in turn are based on what we know of the previous evidence and also on a very wide range of partly remembered experience. Our object is to see to what extent the immediate evidence should modify our preconceptions, or, in other words, to test our hypotheses.

II. INFERENCE

10. Prima facie and Background Evidence.

When we turn from description to inference what can be the general line of our argument? Part of our evidence will be concerned with variations in school practices, such as methods of teaching, organization, and the training of teachers. A second part will be concerned with school outcomes, such as the pupils' achievements and attitudes. A third part will be concerned with the backgrounds against which the practices operate to produce the outcomes. We have chosen our variables to represent these practices, outcomes and backgrounds because our general experience leads us to believe that some practices may produce better outcomes than others. These beliefs are embodied in our hypotheses, and our object is to see whether our evidence supports the hypotheses, and, if so, how strongly. We can begin, in any instance, by seeing whether there is a prima facie case. This will be so if there is a positive association between the outcome variable and the school practice variable. If there is no such association, or if the association is negative, the prima facie evidence tends to confute our hypothesis and we must ask whether the hypothesis is mistaken or whether the evidence in its favor is masked by some other factors. We should not lightly abandom beliefs to which our general experience has led us. But equally fairness demands that we should scrutinize the instances where the prima facie evidence is favorable to our beliefs, to make sure that it is not delusive. In both kinds of instance the consideration of the prima facie evidence is only the first step. We must go on to consider the background evidence. Thus if we are concerned with boys and girls aged 14 we must remember that much has already happened to these children before they come within our view. For example they have been conceived, born, and brought up by their parents in certain neighborhoods, and they have undergone primary and two or three years of secondary schooling. In any particular instance all, or some, of these events may be associated with the school practice, or the outcome, or with both, and taking account of them may change the strength, or even the direction of the prima facie evidence.

11. Surrogates for Past Events.

Ideally we ought to take account of all these events, but plainly we can only do so to the extent that they are represented in our evidence. Generally speaking they are not directly represented, but some of them are represented indirectly by surrogates, such as the present parental occupation and the kind of neighborhood in which the family now lives. Others raise difficult questions. For example it is now firm evidence that at conception a child receives a certain genetic endowment, which may be one of the most important factors governing his future life. Is it fair to say that in a country with a selective system of secondary education the type of secondary school is to some extent a surrogate for genetic endowment, as well as for primary schooling? And in a country where secondary education is not selective are there other variables, such as the pupils stream or track or type of course that can play the same part? Or is this argument circular? To what extent is it legitimate to use one test as a surrogate for background in evaluating outcome measured by another test, on the grounds that the first test represents mainly the effects of background while the second represents mainly the effects of schooling? If we do not use surrogates for background we run the risk of attributing the schooling outcomes that are the effect of background. But if we do use surrogates they are likely to be either too weak, because they fail to cover much of the background, or too strong because they include some foreground as well. Perhaps we should use both weak and strong surrogates, to produce upper and lower estimates of the effects of schooling. Thus a possible surrogate for background, particularly for the foreign language test, is the reading test in the native language, which certainly represents background to a large extent, But it may be too strong because it also includes a good deal of foreground.

12. Cause or Effect?

Another difficulty arises about relations of outcomes of different kinds. Do attitudes determine achievement, or does achievement determine attitudes? Or is the relation reciprocal? These are difficult questions, which need the combined wisdom of our members, both in the choice of variables before the analyses are carried out and afterwards in the interpretation of the results.

13. Regression as Analysis of Variance.

Let us now turn to the simpler question of how the analyses should be carried out when the choice of variables has been made. Our general principle will be to analyze the variation of the outcome, to see how much of it we can associate with the school practice variables, how much with the background variables, and how much we cannot account for because it is associated with events for which we have neither direct measurements nor even surrogates, or which the variation associated jointly with the predictors is divided into successive increments associated with them individually, one must produce arguments not only to justify the choice of predictors but also to justify the order in which they are taken. As a rule, though not invariably, the earlier a predictor is taken the larger its incremental effect. If a predictor has a simple correlation r with the outcome, and a regression coefficient b, in standard measure, when it is taken as one of a particular set of predictors, then the incremental estimates that can be made for it include r^2, when it is taken first in the regression, b^2 $(1-P^2)$ when it is taken last, and rb when it is taken in all positions and the results averaged. P is its multiple correlation with all the other predictors, and the rb estimate is only rough.

14. Regression as hypothesis testing.

Which of these estimates is to be preferred? If our object is to test the hypothesis, that this predictor affects the outcome, the most appropriate estimate is b^2 $(1-P^2)$ which arises when this predictor is taken last. We begin with the prima facie estimate r^2, and note the reduction produced as each background variable is brought in, until finally only b^2 $(1-P^2)$ is left. This is like setting the hypothesis up as a coconut, and hurling at it all the facts we think may be relevant, to see whether we can knock it down. If it survives the despatch of these missiles our confidence in it is much increased.

14.1 It has been objected that if the school practice variables are taken last in this way their effects usually appear to be very small, and that they have been unfairly treated. Certainly this would be the case if other variables were assessed on an earlier position in the regression, unless this position could be justified. If it cannot be justified then the effects of all variables should be reckoned alike, by putting them in the last position. If the variables are well chosen P will usually be fairly small, so that reckoning all variables in the last position will be practically the same thing as reckoning them by their regression coefficients themselves.

15. Some Illustrations.

These points can be illustrated briefly by a few details from the Plowden national survey of primary school children in England (1964). They relate to the analysis within schools for top junior boys (i.e., boys aged 11). There were 17 variables in the regression, besides the achievement of criterion variable. Of the 17 seven emerged with regression coefficients that were more than twice their standard errors. These were:-

1. Parental interest in the child's school work.
2. Parental aspiration for the child.
3. The literacy of the home.

4. The teaching skill, of the child's teacher, as assessed by H.M.I.
5. The teacher's degree of responsibility (graded post).
6. The teacher's in-service training.
7. The teacher's marital status.

Of these the first three, constituting the parental group, were measured in interviews by the Government Social Survey. Other variables in this group, which failed to emerge with significant coefficients, were the father's occupational group, the father's education, the mother's education, the number of dependent children, and the physical amenities of the home. For our present purpose the main interest is to see how the parental and teaching variables compare under the different modes of assessment outlined above. We have:-

	r	b	$b^2(1-P^2)$	rb
Interest	.415	.252	.0536	.1046
Aspiration	.408	.251	.0532	.1024
Literacy	.355	.148	.0177	.0525
Total (Parents)			.1245	.2595
Teaching Skill	.312	.248	.0470	.0774
Teaching Responsibility	.370	.200	.0353	.0740
In Service Training	.067	.150	.0187	.0100
Marital Status	.076	.102	.0099	.0078
Total (Teachers)			.1109	.1692
Total (Parents and Teachers)			.2354	.4287

The two rb assessments add up to .4287, which is R^2. When the variables are put last in the regression the total for teaching is .1109, and for parents .1245. These add to .2354, leaving .1933 on the joint account to make up the .4287. If we take the parents first we find that the teachers are left with 27% of the total. The rb assessments gives teachers 40%. The b^2 $(1-P^2)$ assessment gives the teachers 47%, while if the teachers are taken first they absorb 67%.

15.1 These accounts are not inconsistent. They are answers to different questions. If we take the view that parents should have priority because the Parental influence operates from birth and the influence of these teachers (as opposed to all teachers) only for one year, we ask "What happens when the parents are taken first?" If we could find any reason for supposing that the teachers should have priority we should ask "What happens when the teachers are taken first?" If we adopt neither of these views we ask what happens when teachers and parents are treated alike. In short, the answer obtained depends upon the question asked.

15.2 These are some other interesting points about the little table above. For example, for In Service Training the regression coefficient is more than twice the simple correlation. This is mainly because there is a fairly high negative correlation (-.392) between the measure of teaching skill and the amount of in service training, so that when the measure of skill is held constant the partial correlation between in service training and achievement becomes .216. This is interesting technically because it shows that one cannot safely write off a variable merely because its simple correlation is low. As subject matter it is interesting because it supports other evidence that in service training tends, naturally enough, to be given to weaker teachers. On the first point we may note that the father's occupational group, the father's education, the mother's education, the number of dependent children, and the physical amenities of the home, together with the length of the teacher's experience, all had simple correlations two or three times that for in service training. (The correlation for the number of dependent children was, as would be expected, negative).

15.3 The six variables mentioned at the end of the previous paragraph all fell below the rather low level of significance (F = 3) that was used. On the one hand this meant that they were excluded from the regression, and, on the other hand, that had they been included their coefficients would all have been below .07. Their inclusion would have made little difference to the final value of R^2, or to the regression coefficients of the other variables. But if they had not only been included but also given priority in an incremental account they would have made a considerable contribution to R^2. If we revert to the correlation matrix we find that the five variables describing parental circumstances would, if given priority, account for 13% of the variance, which comes mainly from the difference between the 43% and the 24% given at the foot of the right hand columns in the preceeding table. In other words we can only write them off as negligible if we decide that they do not deserve priority.

16. Summary:

Let us now recapitulate. The cardinal point seems to be one so simple that it is easily taken for granted without enough consideration of its implication. It is that the answers we get from our analyses depend upon the questions we ask. These questions govern both our choice of variables and the order in which we take them. Once choice and order have been determined the analysis extracts the answers from the evidence. But there is nothing in the immediate evidence itself to tell us what the choice and order ought to be. For this we must depend upon a much wider range of partly remembered experience, which makes us give greater credence to some views than to others about what the appropriate model, or in other words the underlying system of causation, may be. Our experiences differ, and our memories are highly selective, so that we can hardly hope to reach unanimity. None the less, if we strive to be explicit we may succeed in crystallizing our differences into two or three alternative models, which could then be offered for the reader's inspection.

16.1 Examples of difficulties in the choice of variables have been given above. Perhaps the major problem is the question of weak and strong surrogates for background. When the choice has been made, r,b, and b^2 $(1-P^2)$ are independent of the order. By tabulating them we can display all the information the analysis can give us until we decide on an order. If we can collect the variables into a small number of homogeneous groups we can display the effect of taking the groups in all possible orders in a compact table such as the one below, where 26 variables have been collected into four groups, and the group contributions are displayed for the 24 possible orders.

Table from IEA/B/40 Appendix I, page 4

2 Variables in the regression

Variable	$\frac{b^2}{c}$	Parts
A	.194	A + AC
B	.012	B + BC
J	.324	AB+ ABC
R	.470	U + C
T	1.000	

Variable	$\frac{b^2}{c}$	Parts
B	.218	B + AB
C	.279	C + AC
J	.118	BC+ ABC
R	.385	U + A
T	1.000	

Variable	$\frac{b^2}{c}$	Parts
C	.143	C + CB
A	.264	A + AB
J	.254	AB+ ABC
R	.339	U + B
T	1.000	

1 Variable in the regression

Variable	$\frac{b^2}{c}$ $(=r^2)$	Parts
A	.518	A+AB+AC+ABC
R	.482	U+B+C+BC
T	1.000	

Variable	$\frac{b^2}{c}$ $(=r^2)$	Parts
B	.336	B+BC+BA+ABC
R	.664	U+A+C+AC
T	1.000	

Variable	$\frac{b^2}{c}$ $(=r^2)$	Parts
C	.397	C+AC+BC+ABC
R	.603	U+A+B+AB
T	1.000	

$$1.000 = A + B + C + AB + BC + CA + ABC$$

A	.072
B	.026
C	.157
AB	.192
BC	-.014
CA	.122
ABC	.132
U	.313
Total	1.000

The accompanying text would give the arguments for the different orders, and perhaps rule some of them out on such grounds as that while the events of early childhood may affect the results of subsequent schooling the reverse effect is not possible.

16.2 By displaying our evidence compactly in this way we could indicate the conclusions to which it led on diverse assumptions, indicate also our own preferences among these assumptions, and enable the reader to draw his own conclusions if his preferences differed from ours.

16.3 In some parts of the report, and notably in the part dealing with Literature, other types of analysis will also be needed. But the basic problems will be the same. In the first place the evidence has to be reduced to a shape compact enough to be grasped. Secondly the assumptions used for drawing inferences from the evidence need to be stated as clearly as possible, so that the critical reader shall not mistake acts of faith for confusions of thought.